Open Research Online

The Open University's repository of research publications
and other research outputs

Comparative and Functional Genomic Analysis of *Streptococcus Equi* and *Streptococcus Zooepidemicus* : Identifying Novel Vaccine Targets

Thesis

How to cite:

Heather, Zoe (2009). Comparative and Functional Genomic Analysis of Streptococcus Equi and Streptococcus Zooepidemicus : Identifying Novel Vaccine Targets. PhD thesis The Open University.

For guidance on citations see FAQs.

© 2009 The Author

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.21954/ou.ro.0000d627

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

COMPARATIVE AND FUNCTIONAL GENOMIC ANALYSIS OF *STREPTOCOCCUS EQUI* AND *STREPTOCOCCUS ZOOEPIDEMICUS*: IDENTIFYING NOVEL VACCINE TARGETS

Zoe Heather BSc (Hons)

A thesis submitted to the Open University for the degree of Doctor of Philosophy in the discipline of Life and Biomolecular Sciences
April 2009

ANIMAL HEALTH TRUST

Abstract

Streptococcus equi subspecies *equi* (*S. equi*) is a host-restricted pathogen of horses and the aetiological agent of strangles. Available evidence suggests that *S. equi* evolved from *Streptococcus equi* subspecies *zooepidemicus* (*S. zooepidemicus*), a versatile bacterium that is often isolated from the equine respiratory tract, but can cause opportunistic disease in horses and other animals. A comparison of the genomes of *S. equi* 4047 and *S. zooepidemicus* H70 and the screening of diverse *S. equi* and *S. zooepidemicus* strains uncovered the genetic events that have shaped the evolution of *S. equi*, and led to its emergence as a niche-adapted pathogen. This analysis provides evidence of functional loss, changes in the organisation and sequence of genes, and pathogenic specialisation through the acquisition of prophage encoding a phospholipase A_2 toxin, and 4 superantigens, and an integrative conjugative element carrying a novel siderophore-like nonribosomal peptide synthetase (NRPS) system. The NRPS shares similarity with the yersiniabactin system found in the high pathogenicity island of *Yersinia pestis* and is the first of its kind to be identified in streptococci. As this genetic feature is absent from the *S. zooepidemicus* population, its gain is considered to have been a key event in the emergence of *S. equi*. Further work determined a role for the NRPS in iron acquisition, and through its heterologous reconstitution in *Escherichia coli* and/or the analysis of allelic replacement mutants in *S. equi*, identified biosynthetic genes, transporters involved in efflux and import of the NRPS product(s), salicylate as a substrate for the NRPS and its regulation by a novel iron-dependent IdeR-like repressor, using various *in vitro* growth assays, including sensitivity to streptonigrin and ^{55}Fe accumulation. Possible vaccine targets were identified in both subspecies and existing diagnostic tools were improved, which included the development of a quantitative PCR test for the detection of *S. equi*.

Acknowledgements

I am particularly grateful to Andrew Waller, my director of studies who has offered endless support during the supervision of my PhD studies. In addition, I received very useful feedback from Nick Davis-Poynter (my supervisor in the early years), Josh Slater (external supervisor), Carl Robinson, Duncan Maskell and my mentor Mark Vaudin. I am thankful to Julian Parkhill who gave me the opportunity to spend 2 months at the Sanger Institute and conduct the comparative genomic analysis. I am particularly appreciative of the endless support provided by Matt Holden since my time at the Sanger Institute. I would like to thank the Animal Health Trust (AHT) for funding these studies and my colleagues at the AHT who supported me throughout and lent me reagents or pieces of equipment. I would particularly like to thank Karen Steward, Fern Ainslie, Melanie Osborne, Debs Flack, Beverley Walladge, Lyndsey Owen, Nicola Butcher and Katy Webb for their help. Karen Steward helped to set up PCR reactions for gene prevalence screening and analysed transcription from a proposed invertible promoter using reverse transcription PCR. Fern Ainslie and Melanie Osborne also contributed to the gene prevalence screening under my supervision. Debs Flack, Beverley Walladge, Lyndsey Owen, Carl Robinson and Andrew Waller all helped with around the clock monitoring during *in vivo* studies and Ken Smith gave fantastic advice on pathology matters. Richard Newton happily provided support with statistical analysis at the drop of a hat. Romain Paillot supplied evidence of superantigen function through mitogenicity assays. Lijiang Song under the guidance of Greg Challis (Warwick University) performed liquid chromatography-mass spectrometry analyses in an attempt to identify the novel metabolite produced nonribosomally by *S. equi*. Greg Challis also gave valuable advice on the bioinformatic analysis of nonribosomal peptide biosynthesis systems. Thanks to Christabel, Katy and Rachael for feeding Kerry when I was stuck in the lab. Last, but by no means least, I must acknowledge the support and endless patience of Brett.

With the exception of assistance acknowledged and sources quoted, all work presented in this thesis is my own work and has not been submitted for any other degree or qualification. Much of the data presented has been published in 2 peer-reviewed scientific articles as indicated in the text.

This thesis is dedicated to my dear beloved Kerry who first introduced me to strangles

Preface

Streptococcus equi subspecies *equi* (*S. equi*) is a highly specialised host-restricted bacterial pathogen and the aetiological agent of strangles, one of the most prevalent infectious diseases of horses worldwide. Strangles is best known for the painful abscessation associated with the disease, which has a major impact on equine welfare. In addition, it is responsible for significant economic loss to the equine industry. Evidence points toward the evolution of *S. equi* from an ancestral strain of *Streptococcus equi* subspecies *zooepidemicus* (*S. zooepidemicus*). However, unlike *S. equi*, *S. zooepidemicus* is generally considered as an opportunistic pathogen. The diverse number of host environments in which *S. zooepidemicus* colonises or causes disease is reflected in the diversity of its genetic make-up.

It was not until 1887 that a distinction between streptococci responsible for strangles and other infections was recognised (Evans, 1936). Since then research has contributed a much better understanding of the population structure of these veterinary pathogens and improved our knowledge of *S. equi* pathogenesis. Some important virulence determinants of *S. equi* have been revealed often through the identification of factors unique to *S. equi* when compared with *S. zooepidemicus*. Given the gravity of strangles, the main driving force for research in this field has come from a need to better contain, diagnose, treat and prevent this disease. However, *S. zooepidemicus* has also received some attention for its association with inflammatory airway disease (IAD) in racehorses and its zoonotic capabilities, as well its role in a range of other diseases in many different species. A need may arise for vaccines or diagnostic tests to combat *S. zooepidemicus* infections, particularly if certain virulent strains are linked to specific disease types.

My research project exploited the newly available genome sequences of *S. equi* strain 4047 (*Se*4047) and *S. zooepidemicus* strain H70 (*Sz*H70). The main aim of the study was to

identify the genetic differences that delineate these two closely related subspecies and to uncover the molecular basis for the evolution of *S. equi* into a specialised pathogen from an opportunistic and generally less pathogenic ancestor. This analysis has the applied objective of identifying novel virulence factors and components for the development of diagnostic tests and vaccines.

Each chapter (2-5) in this thesis includes an introduction, specific aims, methods, results and discussions, and finishes with conclusions. The introductory chapter introduces the pathogenesis and epidemiology of diseases associated with *S. zooepidemicus* infection in horses, including current knowledge regarding their treatment and prevention. Given the versatile nature of *S. zooepidemicus*, diseases associated with infection by this bacterium in other animals are also highlighted. Chapter 1 then describes strangles, the pathogenesis of *S. equi* and the current state of play with regards to treatments, diagnostics and vaccines. The latter part of chapter 1 outlines the phylogenetic relationship of *S. equi*, which is largely clonal within the *S. zooepidemicus* population.

Chapter 2 introduces the molecular processes that generally shape the emergence of specialised pathogens from more flexible non-pathogenic progenitors. Pre-existing knowledge of virulence factors and genetic determinants that are shared between *S. equi* and *S. zooepidemicus* or that distinguish *S. equi* from its proposed ancestor are highlighted. The core of Chapter 2 presents the genetic diversification of these two subspecies following an alignment of the *Se*4047 and *Sz*H70 genomes, gene prevalence screening to establish the true species-specific disparities and some functional analyses to back-up assigned putative functions.

Chapter 3 focuses on the bioinformatic and functional characterisation of ICE*Se2*, a novel *S. equi*-specific integrative and conjugative element (ICE). The acquisition of ICE*Se2* is

proposed to have been a key event in the speciation of *S. equi* and encodes a novel non-ribosomal peptide synthetase (NRPS) system involved in iron acquisition. A novel cation-dependant repressor EqbA regulates the ICE*Se2* NRPS locus. The analysis of recombinant EqbA by gel mobility shift assays is addressed in the latter part of Chapter 3. Reconstitution of the ICE*Se2* NRPS product 'equibactin' was performed using *Escherichia coli* (Chapter 4) and a new strangles diagnostic PCR (polymerase chain reaction) test based on this locus was developed (Chapter 5). Overall conclusions and suggestions for future work are presented in Chapter 6.

Table of contents

List of tables

List of figures

Abbreviations

A domain	Adenylation domain
ABC	ATP-binding cassette
ACT	Artemis Comparison Tool
AHT	Animal Health Trust
CAMP reaction	Christie-Atkins-Munch-Petersen reaction
CAS	Chrome azurol S
CDM	Chemically defined media
CDMs	Chemically defined media supplemented with 10 μM salicylate
C domain	Condensation domain
CDS	Coding sequence
CFU	Colony forming unit
Cp	Crossing point
CRISPR	Clustered regularly interspaced short palindromic repeat
CWSS	Cell wall sorting signal motif
Cy domain	Heterocyclisation domain
DHB	2,3-dihydroxybenzoate
DIG	Digoxigenin
DTT	Dithiothreiotol
ECM	Extracellular matrix
EMSA	Electrophoretic mobility shift assay
Ess	ESAT-6 secretion system
ET	Electrophoretic type
GAS	Group A streptococci
GBS	Group B streptococci
GP	Guttural pouch
GST	Glutathione S-transferase
HA	Hyaluronate
HDTMA	Hexadecyltrimethylammonium
HGT	Horizontal gene transfer
HPI	High pathogenicity island
HPLC	High performance liquid chromatography
ICE	Integrative and conjugative element
IFP	Interstitial fluid pressure
Ig	Immunoglobulin
IPTG	Isopropyl-β-D-thiogalactopyranoside
IS	Insertion sequence
LB	Luria-Bertani
LC-MS	Liquid chromatography-mass spectrometry
LN	Lymph node
MALT	Mucosa associated lymphoid tissue
MCS	Multiple cloning site
MGE	Mobile genetic element
MHC	Major histocompatibility class
MIC	Minimum inhibitory concentration

MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
MM	Minimal media
MSCRAMM	Microbial surface components recognizing adhesive matrix molecule
NP	Nasopharyngeal
NRPS	Non-ribosomal peptide synthetase
NTA	Nitrilotriacetic acid
OD	Optical density
OriC	Replication origin
PBMC	Peripheral blood mononuclear cell
PCP domain	Peptidyl carrier protein domain
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PMN	Polymorphonuclear neutrophil
PTS	Phosphotransferase system
QPCR	Quantitative PCR
RT	Reverse transcription
S. equi	*Streptococcus equi* subspecies *equi*
S. zooepidemicus	*Streptococcus equi* subspecies *zooepidemicus*
Scl	Streptococcal collagen-like protein
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
*Se*4047	*S. equi* strain 4047
SEM	Standard error mean
ST	Sequence type
*Sz*H70	*S. zooepidemicus* strain H70
*Sz*MGCS10565	*S. zooepidemicus* strain MGCS10565
TE	Thioesterase
THA	Todd Hewitt agar
THB	Todd Hewitt broth
THBfcs	Todd Hewitt broth supplemented with 10% foetal calf serum
TMH	Transmembrane helix
TW	Tracheal wash
UDP	Uridine diphosphate
V region	Variable N-terminal region of a surface protein
W region	Proline-rich putative wall spanning region of a surface protein
α_2M	α_2-macroglobulin

Chapter 1 General introduction

1.1 *S. zooepidemicus* infection in horses: clinical disease, pathogenesis and epidemiology

S. zooepidemicus is generally considered a commensal organism of horses and is often isolated from the skin, upper respiratory mucosa, tonsil and associated pharyngeal lymphoid tissues of healthy individuals (Woolcock, 1975). However, *S. zooepidemicus* can cause opportunistic disease given the appropriate conditions, which although poorly understood, are likely to be multifactorial and involve changes or differences in both host and bacterium. A molecular basis for differences in pathogenicity between *S. zooepidemicus* isolates has not yet been reported, but situations arise that appear to render the host more susceptible to disease associated with *S. zooepidemicus* infection. For example, stresses that can occur following training, virus infection, transportation, overcrowding, debilitation, anaesthesia, tissue injury and poor air hygiene have been implicated as a prelude to *S. zooepidemicus* infection. As a result, *S. zooepidemicus* is a common cause of wound infection (cellulitis and abscesses), conjunctivitis and ulcerative keratitis (Brooks *et al.*, 2000), and is the most frequently isolated pyogen from equine joints, lymph nodes (LNs), nasal cavities, and lungs (Hoffman *et al.*, 1993; McGorum, 2006; Timoney, 2004). In mares, cervicitis and endometritis has been associated with *S. zooepidemicus* infection. This might be a sequel to vulvar deformity or to injury during parturition. Abortion can occur when *S. zooepidemicus* is carried in the bloodstream to the placenta or enters the uterus from the distal genital tract (Giles *et al.*, 1993; Hong *et al.*, 1993; Smith *et al.*, 2003; Timoney, 1988).

Young horses are particularly susceptible to opportunistic disease associated with *S. zooepidemicus* infection probably as a result of inadequate immunity, including poor

maternal transfer of antibodies. *S. zooepidemicus* is a common cause of septicaemia in foals and can invade the umbilical stump, causing inflammation of the umbilical veins (omphalophlebitis), bacteraemia and polyarthritis (Timoney, 1988). Bronchopneumonia in foals and pneumonia/pleuropneumonia (including equine shipping fever) in adults, have been associated with mixed infections, although *S. zooepidemicus* is the bacterium most frequently isolated (McGorum, 2006; Raphel & Beech, 1982; Sweeney *et al.*, 1985; Sweeney *et al.*, 1991; Welsh, 1984; Yoshikawa *et al.*, 2003). The disease progresses rapidly and is associated with extensive cellular damage (epithelial desquamation, necrosis), haemorrhage, interstitial changes, and pulmonary consolidation (McGorum, 2006; Oikawa *et al.*, 1994; Yoshikawa *et al.*, 2003). Consequently, although cases of pneumonia are sporadic and have low morbidity, they have a high mortality rate (McGorum, 2006).

Young racehorses also appear more prone to inflammatory airway disease (IAD) (Newton *et al.*, 2003), although time spent in training (and exposure to the new training environment) may be more important than age in determining susceptibility to this condition (Cardwell, 2007). Suffice it to say, IAD is a common problem faced by racehorses early on in their training and has been referred to as a clinical or subclinical lower respiratory tract disease characterised by endoscopically visible tracheal mucus and cytologically evident neutrophilic inflammation (Newton *et al.*, 2003). Poor performance, exercise intolerance and/or coughing are frequent manifestations of IAD. Because respiratory disease has a major impact on racehorses in training (Bailey *et al.*, 1999; Jeffcott *et al.*, 1982; Rossdale *et al.*, 1985) and is frequently publicly attributed as the cause of prolonged periods of poor overall racing performance in yards (Newton *et al.*, 2003), *S. zooepidemicus* has received particular attention for its role in IAD. Epidemiological studies have shown statistically significant associations between infection with *S. zooepidemicus* (and other bacterial species) of the lower airways and both IAD (Burrell *et*

al., 1996; Cardwell, 2007; Chapman *et al.*, 2000; Newton *et al.*, 2003; Wood *et al.*, 1993; Wood *et al.*, 2005a; Wood *et al.*, 2005b; Wood, 1999) and coughing (Christley *et al.*, 2001). Due to the ubiquitous nature of *S. zooepidemicus* and the non-availability of specific-pathogen-free foals to this organism, there is a lack of evidence for *S. zooepidemicus* satisfying Koch's postulates of infectious causality (Newton, 2002). However, 7 out of 9 criteria for causal association of *S. zooepidemicus* with IAD have been met, namely strength of association, consistency, biological gradient, biological plausibility, coherence, some limited experimental evidence, and analogy with other species (Cardwell, 2007; Newton *et al.*, 2003; Wood, 1994). In addition, although most studies identified *S. zooepidemicus* as an important, but not absolutely exclusive infectious factor in this disease syndrome (Burrell *et al.*, 1996; Chapman *et al.*, 2000; Newton *et al.*, 2003; Wood *et al.*, 1993; Wood *et al.*, 2005a; Wood *et al.*, 2005b; Wood, 1999), a study in National Hunt racehorses revealed *S. zooepidemicus* as the only bacterial species associated with IAD in multivariable analysis (Cardwell, 2007). With the exception of influenza virus, which has low prevalence in vaccinated populations, virus infection (including equine herpesvirus-1 and -4, equine rhinovirus-1 and equine adenovirus) has not been associated with clinically apparent respiratory disease in young racehorses (Newton *et al.*, 2003). *S. zooepidemicus*-associated IAD is therefore unlikely to occur as strictly secondary to primary viral infection, which has been the traditional view.

It is likely that the aetiology of IAD involves stresses associated with the lifestyle of the racehorse, including those that arise from stabling (Christley *et al.*, 2001). However, lower airway disease has also been observed in recently weaned ponies entering new premises, suggesting that non-Thoroughbreds kept at pasture undergo a form of IAD clinically similar to that seen in racehorses (Newton *et al.*, 2007). This natural clinical respiratory disease has been characterised predominantly by mucoid nasal discharge, coughing and occasionally pyrexia, dyspnoea and pneumonia. *S. zooepidemicus* is one of the bacterial

species most commonly isolated from the lungs of affected animals and an epidemiological study found that higher clinical scores were associated with higher colony numbers of *S. zooepidemicus* in tracheal wash (TW) samples (dose response), which was consistent with previous studies of both clinical disease and airway inflammation (Newton *et al.*, 2007). Interestingly, ponies suffering naturally occurring respiratory disease occasionally present submandibular lymphadenopathy (Newton, 2002; Newton *et al.*, 2007) (unpublished observations). These LNs become enlarged and hardened and may become abscessed, which often occurs during *S. equi* infection, although abscessation associated with *S. zooepidemicus* infection is much less frequent. Unlike pneumonia, less severe IAD occurs as outbreaks with high morbidity and very low mortality, possibly because horses in training or immunologically naive youngsters experience less severe compromise to their respiratory tract defences than sick foals or adults subject to greater stress, such as anaesthesia (McGorum, 2006).

The molecular epidemiology of *S. zooepidemicus* respiratory infection has been studied in detail in 2 different equine populations: Thoroughbred racehorses in training and Welsh mountain ponies with naturally occurring respiratory disease following their recent weaning and transportation to new premises (Barquero *et al.*, 2009; Newton *et al.*, 2008). Both studies typed strains of *S. zooepidemicus* based on the characterisation of the Szp-adhesin hypervariable region (Walker & Timoney, 1998) and the 16S-23S rRNA gene intergenic spacer (Chanter *et al.*, 1997) so that the prevalence of different *S. zooepidemicus* types at different locations in the respiratory tract could be assessed over the study period (10 months in the Thoroughbred study and 10 weeks in the pony study).

S. zooepidemicus infection was highly prevalent during the pony study, being isolated from 94% of TW samples and 88% of nasopharyngeal (NP) swabs (Newton *et al.*, 2008), but was markedly less prevalent in the Thoroughbred study (22 to 23% of TW and NP

samples), which sampled an older population (Barquero *et al.*, 2009). A wide diversity of *S. zooepidemicus* types were identified in both equine populations, with 24 different types isolated from 198 Thoroughbreds and 39 types recovered from 29 ponies, although 10 types accounted for > 80% of all isolates recovered from the 2 populations. Some similarities were seen in the most common types isolated, with A1 types being the most prevalent in both study populations. The *S. zooepidemicus* genome sequencing strain (*Sz*H70) is type A1HV4, which was the 5th most prevalent type isolated from tracheal and NP samples in the pony study (Newton *et al.*, 2008), the most prevalent type isolated from NP samples in the Thoroughbred racehorses and the 2nd most prevalent type isolated from TWs in the same Thoroughbred population (Barquero *et al.*, 2009). However, differences in the occurrence and prevalence of types were apparent between the 2 different study populations and also between different Thoroughbred yards (Barquero *et al.*, 2009; Newton *et al.*, 2008). In both studies, more types were isolated from the trachea (23 to 33) than the nasopharynx (15 to 27) but temporal patterns of bacterial prevalence for the respective 4 most prevalent types were similar in TW and NP samples. These data suggest that, counter to the observations previously reported (Timoney *et al.*, 1997a), the trachea acts as a source of *S. zooepidemicus* for the upper respiratory tract due to the clearance of mucus, bacteria and debris proximally with mucociliary action. Different *S. zooepidemicus* strains may also be adapted to differences in conditions for colonisation between the upper and lower airways (Barquero *et al.*, 2009; Newton *et al.*, 2008). It was noted that in individual animals NP *S. zooepidemicus* types were not necessarily predictive of types isolated from the trachea (Newton *et al.*, 2008). More *S. zooepidemicus* isolates were cultured from TW relative to NP samples, which may also have contributed to more types arising from the former (Barquero *et al.*, 2009; Newton *et al.*, 2008).

Different *S. zooepidemicus* types varied markedly in their prevalence over time with some appearing dominant at times and many others never becoming well established (Newton *et*

al., 2008). The appearance and disappearance of the different types over time in the pony population suggested that there was clonal succession of *S. zooepidemicus* types amongst the group of ponies (Newton *et al.*, 2008). A 3rd study assessed the infection/colonisation patterns of *S. zooepidemicus* strains (typed by multilocus sequence typing (MLST)) in National Hunt racehorses and provided preliminary evidence for recurrent or persistent infection with the same strains, clonal succession within individual animals and sequential new infection with different strains (Cardwell, 2007).

There is emerging evidence that horse genotype plays a role in *S. zooepidemicus* disease susceptibility, possibly related to transferrin haplotype. In the pony study of naturally occurring equine respiratory disease, ponies with transferrin haplotypes D were at significantly decreased risk of disease and *S. zooepidemicus* infection, whereas those with haplotype F2 were at significantly increased risk (Newton *et al.*, 2008). Similarly, in a study of the effect of transferrin haplotype on susceptibility to IAD in young racehorses, individuals with haplotype D had a significantly lower prevalence and duration of IAD compared to those without this haplotype (Newton, 2003). Since transferrin is a host glycoprotein involved in sequestering free iron, these data suggest that the availability of iron to *S. zooepidemicus* during *in vivo* infection may differ depending on host transferrin genotype and impact on the ability of *S. zooepidemicus* to cause disease. Alternatively, another host factor genetically linked to transferrin haplotype may be important in disease susceptibility. Similar trends have been observed in horses infected with other bacterial species (Mousel *et al.*, 2003; Newton, 2003), although a study of IAD in National Hunt racehorses showed that whilst increased odds of tracheal mucus were associated with the possession of haplotype F, the risk of nasal discharge was increased in horses with haplotype D (Cardwell, 2007). The apparently contradictory effect of haplotype D compared with other studies may have been related to the predominance of D/F heterozygotes in the study (Cardwell, 2007).

1.1.1 Immunity, vaccines, diagnostics and treatment

Most horses carry precipitating antibodies to *S. zooepidemicus*, probably as a consequence of NP carriage of the organism and serum antibodies can cross-react with *S. zooepidemicus* antigens shared by *S. equi* (Timoney & Trachman, 1985). Antibodies to the hypervariable adhesin of *S. zooepidemicus* (Szp) have been detected in equine serum (Timoney *et al.*, 1995). During the development of a strangles blood test at the Animal Health Trust (AHT), serum antibodies specific to a number of antigens present in *S. zooepidemicus*, as well as *S. equi*, were present in negative control samples collected from horses with no history of *S. equi* infection (Nicola Butcher, unpublished observations). Another strangles blood test based on the *S. equi* SeM anti-phagocytic surface antigen requires a pre-absorption step with heat killed *S. zooepidemicus* to reduce cross-reactivity from antibodies raised to the *S. zooepidemicus* variant of the SeM protein (Timoney, 1997). These data point to the development of humoral immune responses following *S. zooepidemicus* infection, but these have not been well studied for their role in clearing the bacterium. Antisera raised against Szp in rabbits elicited a protective opsonic activity against *S. zooepidemicus in vitro* (Timoney *et al.*, 1995). In addition, mice vaccinated with an *E. coli* lysate containing recombinant Szp were significantly protected against *S. zooepidemicus* infection following intraperitoneal challenge (Timoney *et al.*, 1995). However, because this antigen of *S. zooepidemicus* is highly variable, immunity may not be cross-protective against other *S. zooepidemicus* strains. Antibodies and complement in uterine fluids have also been shown to be effective in enhancing phagocytosis and clearance of *S. zooepidemicus* (Asbury *et al.*, 1984).

The presence of *S. zooepidemicus* as a normal resident of the equine upper respiratory tract and its regular isolation from adult horses, suggests that this organism is able to persist in

the respiratory tract despite the development of host immunity, which may in part be explained by the diversity of this bacterium. The sequential infection with different types of *S. zooepidemicus* previously observed (described above) (Cardwell, 2007; Newton *et al.*, 2008), suggests that continuous *S. zooepidemicus* infection does not necessarily reflect continuous colonisation by a single type, but instead may be due to clonal succession with non-cross-immunising subtypes of *S. zooepidemicus*. In support of this there is evidence that equine sera can be highly discriminating in their ability to opsonise *S. zooepidemicus* isolates (Causey *et al.*, 1995). The predisposition of young horses to *S. zooepidemicus* associated disease (described above) indicates the importance of immunity, probably to a wide range of subtypes, in preventing these opportunistic infections.

No vaccines are currently available to protect against diseases associated with *S. zooepidemicus* infection and the development of an efficacious vaccine is likely to be hampered by the diversity of this subspecies. An attempt was made to protect recently weaned ponies from natural respiratory tract infection with *S. zooepidemicus* using inactivated, alum adjuvanted vaccine containing 2 strains each of *S. zooepidemicus* and *Actinobacillus equuli* (Newton *et al.*, 2007). However, this vaccine failed maybe in part due to failure of challenge by the specific subtypes of the bacteria in the vaccine combined with an absence of cross-protective immunity conveyed between subtypes and an overwhelming mixed infection challenge (Newton *et al.*, 2007). Future vaccine development may require the targeting of subtypes and conserved virulence factors linked to specific disease types. Immunisation of mares with bacterial protein extracts provided some resistance to endometritis caused by *S. zooepidemicus*, although vaccination resulted in adverse reactions at the injection site in some horses (Widders *et al.*, 1995).

Culture analysis from relevant swab or wash samples can be used to identify *S. zooepidemicus*. This relies on the isolation of β-haemolytic streptococci that agglutinate

latex particles coated with antibodies to Lancefield group C polysaccharide cell wall antigen and ferment sorbitol, lactose, ribose (most strains) but not trehalose (Bannister *et al.*, 1985). A multiplex PCR test, based on a *S. equi* subspecies-specific region of the superoxide dismutase gene *sodA* and the *S. equi*-specific mitogenic toxin gene *seeI*, has been developed with the potential to detect *S. zooepidemicus* and distinguish this bacterium from *S. equi* in test samples (Alber *et al.*, 2004). However, PCR testing is not commonly used to diagnose *S. zooepidemicus* infections at present, although Zoologix is currently marketing a qualitative PCR test to identify *S. zooepidemicus* as a tool in the detection and prevention of haemorrhagic pneumonia in dogs (see below) (http://www.zoologix.com/dogcat/Datasheets/StreptococcusZooepidemicus.htm, accessed 23.06.09).

S. zooepidemicus is generally sensitive to penicillin and related antibiotics and also a variety of other antibiotics including tetracyclines, trimethoprim-sulfonamide combinations and peptide antibiotics (McGorum, 2006). Antibiotics have been used to treat *S. zooepidemicus* infections and also as a preventive measure to predisposing medical conditions or stressors (e.g. uterine fluid accumulation). Limiting stress through suitable management practices is equally important in preventing opportunistic disease.

1.2 Diseases associated with *S. zooepidemicus* infection in other animals

Unlike *S. equi*, *S. zooepidemicus* is not restricted to the horse and is an opportunistic pathogen of many other animal hosts including cattle, sheep, goats, foxes, birds, rabbits, guinea pigs, dogs, pigs and monkeys (Chalker *et al.*, 2003; Pesavento *et al.*, 2008; Stableforth, 1959). Associated diseases include mastitis, polyarthritis, bronchopneumonia, pleuritis, epicarditis, endocarditis, and meningitis (Las Heras *et al.*, 2002; Salasia *et al.*,

2004; Sharp *et al.*, 1995; Soedarmanto *et al.*, 1996; Stevenson, 1974). Distinct strains of *S. zooepidemicus* appear to be associated with certain disease types (Webb *et al.*, 2008). Related *S. zooepidemicus* strains have been isolated from recent outbreaks of fatal haemorrhagic pneumonia in dogs in the USA and the UK (Webb *et al.*, 2008). A single virulent clone appears to be associated with a disease outbreak in 1994 in Indonesia among pigs and monkeys (Soedarmanto *et al.*, 1996). On rare occasions, *S. zooepidemicus* can cause invasive infections in humans including bacteraemia and meningitis. These infections usually originate following zoonotic transmission from domesticated animals to humans (Barnham *et al.*, 1987; Bradley *et al.*, 1991; Downar *et al.*, 2001; Hashikawa *et al.*, 2004; Korman *et al.*, 2004; Latorre *et al.*, 1993; Pati *et al.*, 2007). Epidemic outbreaks have been associated with the consumption of unpasteurised dairy products, and many of these episodes have been complicated with post-streptococcal glomerulonephritis (Bordes-Benitez *et al.*, 2006; Francis *et al.*, 1993; Kuusi *et al.*, 2006), also a sequela of *S. zooepidemicus* infection in horses (Divers *et al.*, 1992). Immune reactant renal disease has been associated with IgG antibody and streptococcal antigens (Divers *et al.*, 1992).

1.3 Strangles

Strangles is one of the oldest recorded equine diseases. Giordano Ruffo, farrier to Emperor Frederick II of Italy, and Albertus Magnus, a Dominican bishop, described the acute clinical disease strangles (*strangulina*) in the earliest surviving European veterinary manuscripts dated from the 13th century (Slater, 2003). Caused by infection with the bacterium *S. equi*, strangles is a disease of the upper respiratory tract and draining LNs. Abrupt onset of fever is followed by acute swelling with subsequent abscess formation in the submandibular and retropharyngeal LNs and sometimes other LNs of the rostral neck (parotid and cranial cervical). This lymphadenopathy is a major clinical sign of the disease. Mature LN abscesses rupture, internally or externally, to release sometimes profuse quantities of creamy pus. Early on in the disease, pharyngitis and laryngitis may

contribute to serous nasal secretions, whilst later on the internal drainage of a ruptured abscess is characterised by a mucopurulent nasal discharge. Ocular discharge can also occur. The retropharyngeal LNs may drain into and cause empyema of the guttural pouch (GP). Throughout the course of the infection horses may demonstrate dysphagia and general depression as a result of fever and/or pain from inflammation of the respiratory tract, the development of abscesses and GP empyema. A soft moist cough sometimes accompanies these manifestations. Complications and severe sequelae to strangles infection often result in fatality - an 8% fatality rate was reported in one study (Sweeney *et al.*, 1987).

1.3.1 Pathogenesis

S. equi enters via the nose and mouth and attaches to the mucosa associated lymphoid tissues (MALT) of the oro- and nasopharynx. There is no evidence for colonisation prior to invasion. After infection, the bacterium is difficult to detect on the mucosal surface but small numbers appear in the tonsillar crypts, in the adjacent follicular subepithelium and underlying lymphoid follicles (Sweeney *et al.*, 2005; Timoney & Kumar, 2008). Translocation occurs in a few hours to the efferent LNs that drain the pharyngeal and tonsillar region. Intracellular and extracellular multiplication of *S. equi* has been detected in the tonsillar lymphoid tissue and LNs (Timoney & Kumar, 2008). Once inside the LNs, large numbers of polymorphonuclear neutrophils (PMNs) are attracted by complement-derived chemotactic factors, but fail to phagocytose and kill the streptococci. Consequently, extracellular streptococci accumulate in long chains surrounded by large numbers of degenerating neutrophils (Sweeney *et al.*, 2005; Timoney & Kumar, 2008). Final disposal of these organisms is dependent on lysis of the abscess capsule and evacuation of its contents, which usually occurs within 7 to 14 days after the first clinical signs of disease. Fever develops between 3 and 14 days after exposure and blood

fibrinogen concentrations, white blood cell counts and neutrophil counts increase. The progression of strangles and severity of disease depends on the dose of *S. equi* received (the larger the dose, the shorter the incubation period) and depends on the immune status of the horse (e.g. more severe disease may be seen in young and elderly horses) (Sweeney *et al.*, 2005; Timoney & Kumar, 2008).

Severe LN enlargement may lead to complications. The name strangles was coined because affected horses can be suffocated by enlarged LNs that caused dyspnea. This may occur through restriction of the airway (compression of the pharynx, larynx or trachea) or may result from damage to laryngeal nerves (Sweeney *et al.*, 2005). Resultant respiratory distress and/or dysfunction of the swallow reflex as a result of nerve damage (Jeremy Kemp-Symonds, personal communication) can cause aspiration pneumonia (Sweeney *et al.*, 1987; Sweeney *et al.*, 2005).

Metastasis of *S. equi* can result in abscess formation in other locations such as the lungs, brain, thoracic or abdominal LNs. This condition referred to with the archaic term 'bastard strangles' is almost always fatal. Spread is presumed to occur mainly via lymphatic channels as bacteraemia is rare (Evers, 1968) and LNs are usually involved alongside solid organs in the disseminated disease (e.g. splenic LNs and the spleen or mesenteric LNs and the intestine) (Ken Smith, personal communication). However, haematogenous dissemination of antigens may contribute to the development of purpura haemorrhagica, another sequela of strangles.

Purpura haemorrhagica (also triggered by other streptococcal infections) is an immune-system mediated aseptic vasculitis characterised by subcutaneous oedema, which is frequently fatal. Interestingly, this condition is also associated with the vaccination of horses with strangles vaccines containing the antiphagocytic *S. equi* surface protein SeM

(Pusterla *et al.*, 2003). Affected horses have characteristically very high levels of anti-SeM-protein antibodies and it is thought that these generate SeM-protein immune complexes that deposit in blood vessel walls to cause a type 3 hypersensitivity reaction (Galan & Timoney, 1985b). An alternative mechanism for this state has been proposed based on the activation of host neutrophils by the binding of M-protein-fibrinogen complexes to β_2 integrin receptors (Waller, 2007). This occurs in *Streptococcus pyogenes* mediated streptococcal toxic shock syndrome in humans and results in the release of an inflammatory mediator (heparin binding protein) by activated neutrophils with subsequent vascular leakage, oedema and toxic shock. In a mouse model of *S. pyogenes* infection, severe pulmonary damage caused by vascular leakage was substantially reduced by treatment with a β_2 integrin antagonist (Herwald *et al.*, 2004).

1.3.2 Transmission and persistent carriage

Discharge from nasal secretions and ruptured abscesses provides a source of *S. equi* that can be transmitted directly to other horses through normal social behaviour or indirectly through shared housing, water and feed sources, and other equipment or the clothing of handlers, farriers or veterinarians. Transmission may occur when horses show no obvious clinical signs of disease, for example during the early infection phase or after the clinical signs have disappeared. A moderate proportion of horses continue to harbour *S. equi* for several weeks after clinical signs have gone, even though the organism is no longer detectable in the majority 4 to 6 weeks after apparent recovery (Sweeney *et al.*, 2005).

Incomplete drainage of exudates from the GPs (and more rarely the sinuses), which are known to become infected in the early phase of infection and following rupture of the adjacent retropharyngeal LNs, contributes to the prolonged intermittent shedding of *S. equi* (Newton *et al.*, 1997; Sweeney *et al.*, 2005). Drying and hardening of exudates leads to

the formation of discrete bodies called chondroids that can remain in the GP for several years. Consequently, up to 10% of horses that have apparently recovered from strangles can continue to harbor *S. equi* for many months or years (Newton *et al.*, 1997) and transmit the infection to naïve horses. The existence of carriers is undoubtedly important (Fintl *et al.*, 2000; Judy *et al.*, 1999; Newton *et al.*, 1997; Newton *et al.*, 2000; Timoney *et al.*, 1998) and probably critical to the success of *S. equi.*

S. equi does not readily survive in the presence of soil-borne microflora and prolonged environmental persistence of *S. equi* is unproven (Sweeney *et al.*, 2005). One laboratory-based study demonstrated survival on wood for 63 days at 2°C but did not include co-infection with common environmental bacteria (Jorm, 1992). The inability of *S. equi* to persist in the environment suggests that carriers play a key role in the recurrence of strangles outbreaks.

1.3.3 Diagnosis, control and treatment

Careful management, treatment and diagnostic testing enable *S. equi* transmission to be contained and can prevent the initiation of new outbreaks. Horses that are new to premises or are suspected of being infectious and any individuals that they have been in contact with can be kept in isolation until further tests and/or time reveals the risk they pose to others. Monitoring body temperature can identify the onset of fever, an early clinical sign of strangles. Prompt treatment with antibiotics during the early acute phase may abort the infection but does not prevent re-infection if the animal is re-exposed once the treatment is withdrawn (Sweeney *et al.*, 2005). *S. equi* is sensitive to the majority of antibiotics *in vitro*, but antibiotic therapy is usually ineffective once external lymphadenopathy is detected probably due to the lack of sufficient vascularity in the abscess to enable antibiotic penetration to therapeutic levels (Harrington *et al.*, 2002). Management is then

limited to isolation, careful monitoring (particularly for complications associated with strangles) and palliative care until the abscesses have ruptured and the horse's own immunity has cleared the infection.

PCR and culture analysis of NP swabs or wash samples are useful tools to identify *S. equi* infection, particularly where clinical signs of disease are not obvious (e.g. during the early and late phases of disease, mild disease or asymptomatic carriage) (Newton *et al.*, 2000; Sweeney *et al.*, 2005; Timoney & Artiushin, 1997). Culture analysis relies on the isolation of β-haemolytic streptococci that agglutinate latex particles coated with antibodies to Lancefield group C polysaccharide cell wall antigen and are unable to ferment trehalose, sorbitol, lactose or ribose (Bannister *et al.*, 1985). Unfortunately the growth of *S. equi* on agar plates can be outcompeted or masked by other bacteria particularly if *S. equi* is present as the minority. PCR can help to overcome this limitation but both tests require *S. equi* to be present in the respiratory tract at the moment of sampling. As *S. equi* shedding is usually intermittent, a prolonged quarantine period with or without repeated testing is often necessary. In the past, an individual has been considered clear of infection on the collection of 3 consecutive negative NP swabs over a 2 week period (George *et al.*, 1983). Currently there is an increasing trend toward blood testing (see below) and GP endoscopy. Endoscopy is used to look for signs of pathology in the GP, to collect lavage samples for analysis by culture/PCR and to clear any chondroids or contaminating material (often followed by antibiotic treatment to kill residual bacteria).

Blood tests bestow a significant advancement in the detection of *S. equi* as they do not require an animal to be shedding *S. equi* at the time of testing. An ELISA test is currently used by practitioners in the USA (http://www.idexx.com/equine/laboratory/sequi_elisa/. Accessed 28.10.08). Unfortunately, overlapping breakpoints in normal and convalescent horses often complicates the interpretation of results from this SeM-protein-based test.

SzM, a near identical partial homologue to SeM is present in *S. zooepidemicus* (Kelly *et al.*, 2006) and antibodies raised to SzM are likely to cross-react with the SeM antigen. Pre-absorption of cross-reacting antibodies by incubation of sera with heat killed *S. zooepidemicus* improves the reliability of the SeM ELISA, although at the expense of assay sensitivity (Timoney, 1997; Waller & Jolley, 2007).

A new ELISA test was recently developed in the UK (http://www.aht.org.uk/bact_blood.html. Accessed 28.10.08), which has a sensitivity of 91.5% and specificity of 90.2% and enables the robust detection of horses that have been recently exposed to *S. equi*. The antigens used in this test were chosen from among the array of coding sequences in the early *Se*4047 genome sequence predicted by bioinformatics to encode proteins likely to be exposed to the host immune system. This current study further revealed the antigens to be a specific feature of *S. equi* and absent from the *S. zooepidemicus* population (SEQ2190 and SeeI, Figure 2.6, Chapter 2).

Blood sampling is generally well tolerated by horses and large numbers of animals can be screened by this method, which has a 24 hour turnaround time. However, antibody levels take 2-3 weeks to build up post infection and can remain elevated for up to 6 months once the horse has recovered and eliminated *S. equi*. The blood test is therefore a suitable pre-screen procedure to complement good management, GP endoscopy, PCR and culture analysis.

1.3.4 Vaccines

Most horses develop immunity during recovery from strangles that may persist for 5 years or longer (Hamlen *et al.*, 1994; Sweeney *et al.*, 2005; Waller & Jolley, 2007). This indicates that stimulation of a high level of immunity is biologically feasible given the

appropriate presentation of protective immunogens. Despite this, progress in the development of an effective strangles vaccine has been slow. Bacterin type vaccines produced by moderate heat inactivation of logarithmic phase cultures were introduced in Australia in the 1940s and then in the USA in the 1960s (Timoney, 2004). Unfortunately, unacceptable severe adverse reactions occurred at the injection site, including inflammation and abscess formation. Extract vaccines prepared by hot acid or mutanolysin plus detergent treatment to release acid-resistant or surface proteins were better tolerated than bacterins and were marketed in the USA in the 1970s and 1980s. However, these conventional vaccines have shown disappointing efficacy, with little published data to support significant protection (Jorm, 1990; Sweeney *et al.*, 2005; Timoney & Eggers, 1985).

More recent progress has seen the application of live attenuated *S. equi* vaccines. The first, Pinnacle IN, is administered via the intranasal route and has been widely used in the USA since its launch in 1998. However, the attenuating mutations in this strain were introduced through chemically-induced random mutagenesis and therefore, except for a resultant lack of capsule production and defects in carbohydrate utilisation, the mutations have not been defined. Point mutations are prone to back mutation and thus to reversion to full virulence and although this vaccine may protect up to 100% of horses (Timoney, 1993; Walker & Timoney, 2002; Waller & Jolley, 2007) it has not been licensed for sale in Europe due to safety concerns. These include residual virulence with formation of slowly developing submandibular abscesses in a small percentage of vaccinates, nasal discharge and occasional cases of purpura haemorrhagica (Sweeney *et al.*, 2005). The Pinnacle strain has since been refined to improve the genetic stability of its acapsular phenotype by deletion of part of the *hasA* and *hasB* genes, which encode hyaluronate (HA) synthase and uridine diphosphate (UDP)-glucose dehydrogenase, respectively (Pinnacle HasNeg) (Walker & Timoney, 2002). Both enzymes are involved in HA capsule biosynthesis. It is, however,

not clear if this new strain has replaced the original vaccine strain in the commercial product (Meeusen *et al.*, 2007) or what this modification offers in terms of improved safety.

In 2004, the Equilis StrepE vaccine, a live recombinant strain TW928 was licensed for use in Europe (Jacobs *et al.*, 2000; Waller & Jolley, 2007). This strain, created by deletion of the *aroA* gene (Kelly *et al.*, 2006), was 10^4-fold attenuated in intraperitoneal mouse challenge studies (Waller & Jolley, 2007). Originally developed for intranasal application, protection was only accomplished in horses by intramuscular injection and an unusual route of subcutaneous administration into the upper lip (Jacobs *et al.*, 2000). As seen with the conventional vaccines, intramuscular inoculation caused undesirable adverse reactions, including local swelling of muscle tissue and the formation of abscesses. Small pustules also occurred at the injection site following subcutaneous inoculation in the lip. However, the scale of the immune response to vaccination and persistence of the attenuated organism at the injection site appeared to correlate with protection. Intramuscular vaccination conferred 100% protection from subsequent *S. equi* challenge, whilst lip inoculation protected 50% of horses from the development of LN abscesses and a further 25% of vaccinates had reduced clinical signs of disease, albeit for a limited (3-month) duration of immunity. On dose reduction, reduced lip reactions further correlated with decreased protection (Jacobs *et al.*, 2000).

To reduce side effects and at the expense of efficacy, Equilis StrepE is administered via the lip route and recommended for use in horses of high or moderate risk of strangles where acquisition of a short duration of immunity is advantageous. Booster vaccination of horses inoculated up to 6 months previously in the face of an outbreak has been shown to improve clinical outcome (Waller & Jolley, 2007). However, bacterial replication of TW928 and abscess formation has been reported in a retropharyngeal LN following vaccination

(Kemp-Symonds *et al.*, 2007) and like the Pinnacle vaccine, contamination during concurrent immunisation with other vaccines has resulted in inadvertent adverse reactions at parenteral injection sites. Equilis StrepE is currently unavailable due to quality control problems and new safe methods of immunisation against strangles are being investigated with the use of novel recombinant strains (Andrew Waller, unpublished data) and subunit vaccines.

Vaccines that target the SeM protein of *S. equi* specifically, showed promise in mouse vaccination challenge studies, but failed to demonstrate significant protection in horses despite the generation of SeM reactive antibodies (Hoffman *et al.*, 1991; Meehan *et al.*, 1998; Sheoran *et al.*, 2002; Timoney & Mukhtar, 1993; Timoney *et al.*, 1997b). Similarly, a recombinant *S. equi* hyaluronate associated protein (HAP) vaccine, was partially protective in mice (Chanter *et al.*, 1999) but failed to prevent the development of strangles in horses despite the generation of serum and mucosal antibodies to this protein (Neil Chanter, unpublished data).

A recent study assessed 2 multi-component subunit vaccines comprising 5 to 6 recombinant surface exposed or secreted *S. equi* proteins that are immunogenic and recognised by the host immune system during experimental infection with *S. equi*. These included cell wall-anchored surface proteins (SzPSe, Cne, Se46.8, Se51.9), a lipoprotein (Se44.2) and a secreted protein (Se42.0) with functions as diverse as fibrinogen binding, collagen binding, adhesion, immunoglobulin (Ig) and platelet binding and methionine sulfoxide reductase activity. Subcutaneous immunisations yielded strong serum antibody responses to each protein component but failed to provide any protection against strangles (Timoney *et al.*, 2007).

Another group has successfully identified a multicomponent recombinant protein vaccine that confers some protection from strangles in horses (Waller *et al.*, 2007) (Table 1.1). This group used a mouse model to assess the suitability of different protein combinations in vaccines to protect against *S. equi* infection (Table 1.1). Their early results demonstrated that a combination of 3 proteins Fnz (cell wall-attached fibronectin binding protein), Sfs (secreted fibronectin binding protein) and Eag (α_2-macroglobulin, albumin and IgG binding protein) conferred improved protection of mice when compared with vaccination with Eag alone. In addition, vaccination with the collagen binding protein Cne or a collagen-like protein SclC conferred partial protection which was enhanced on vaccination with a combination of Cne and Eag (Flock *et al.*, 2006). Picking the right combination of vaccine components to generate a synergistic and protective immune response appears to be important. Intranasal and intramuscular immunisation of horses with a combination of Eag, Cne and SclC resulted in significantly reduced nasal discharge and bacterial growth from NP swabs and a lower rate of empyaema or scarring of the GPs following intranasal challenge of horses with *S. equi* (Waller *et al.*, 2007).

Table 1.1 Potential subunit vaccine components to protect against strangles

Protein	Function	Features	Protection against *S. equi* infection following recombinant protein vaccination
Fnz	CL- and Fn-binding	Surface protein	Partial protection in mice when combined with Sfs and Eag
Sfs	Fn- binding	Secreted protein	Partial protection in mice when combined with Fnz and Eag
Eag	IgG-, α_2M-, albumin-binding	Surface protein	Partial protection in mice enhanced when used in combination Fnz and Sfs, also partial protection in mice when used in combination with Cne
			Partial protection in horses in combination with Cne and SclC
Cne	CL- binding	Surface protein	Partial protection in mice enhanced when used in combination with Eag
			Partial protection in horses in combination with Eag and SclC
SclC	Unknown	CL-like surface protein	Partial protection in mice
			Partial protection in horses in combination with Eag and Cne

CL: collagen, Fn: fibronectin, α_2M: α_2-macroglobulin

1.3.5 Immunity

The basis for protection against *S. equi* infection is not fully understood. Humoral immunity is presumed to be an important aspect of defence as serum antibodies specific to *S. equi* antigens have been observed in horses that are resistant to *S. equi* infection as a result of previous exposure or vaccination (Galan & Timoney, 1985a; Timoney *et al.*, 2007). Convalescent sera contain antibodies raised against a wide range of *S. equi* antigens (Andrew Waller, unpublished data), (Artiushin *et al.*, 2002; Timoney *et al.*, 2008). However, a strong vaccine-induced antibody response often does not correlate with protection (see above), which may depend on the target(s) or type of vaccine chosen or a

failure to elicit a suitable antibody response or as yet unknown mechanisms of immunity (e.g. cell mediated immunity). The exact mechanism of action of antibodies has rarely been addressed but is proposed to depend on the blocking of function and/or opsonisation. For example, Cne mediates bacterial adherence to collagen, a component of host extracellular matrix (ECM) (Lannergard *et al.*, 2003) and serum antibodies against Cne are able to block adherence of *S. equi* to immobilized collagen *in vitro* (Flock *et al.*, 2006). Antibody-mediated blockage of adherence *in vivo* might therefore hamper the infectious process and contribute to the partial protection observed in animals vaccinated with Cne as a component of muti-subunit vaccines (see above). Convalescent ponies or those immunized with recombinant *S. equi* mitogen SeeI (SePE-I) are protected from the pyrogenic effect of subsequent SeeI administration most probably through a neutralising antibody response (Artiushin *et al.*, 2002). Neutralising antibodies to streptococcal superantigens appear important in resistance of humans to severe group A streptococcal disease (Kotb, 1995).

Opsonophagocytosis of *S. equi* has been demonstrated indirectly using *in vitro* bactericidal assays and sera collected following *S. equi* infection or vaccination with M protein, extract or bacterin vaccines (Sheoran *et al.*, 1997; Timoney & Eggers, 1985). Ig subclasses capable of eliciting this effector function have been defined with the aid of genomic identification. IgG1, IgG3, IgG4, IgG5 and IgG7, but not IgG2 and IgG6, are able to elicit a strong respiratory burst from equine peripheral blood leukocytes suggesting that these subclasses are able to interact with Fc receptors on effector cells. IgG1, IgG3, IgG4 and IgG7, but not IgG2, IgG5 and IgG6, are able to bind complement C1q and activate complement via the classical pathway (Lewis *et al.*, 2008b). Of the originally described IgG subclasses, IgGa corresponds to IgG1, IgGb to IgG4 and IgG7, IgGc to IgG6, and IgG(T) to both IgG3 and IgG5 (Wagner *et al.*, 2002; Wagner *et al.*, 2004). Purified IgGa (IgG1) and IgGb (IgG4 and IgG7) but not IgGc (IgG6) or IgG(T) (IgG3 and IgG5) from

sera collected after vaccination with *S. equi* protein extract killed *S. equi in vitro* in the presence of complement and phagocytes (Sheoran *et al.*, 1997).

A protective antibody response might therefore contain a predominance of Ig sub-isotypes that contribute to the clearance of bacteria via these mechanisms and some antigen-specific Ig subclass responses to infection and vaccination have been measured. One study demonstrated a predominant and prolonged serum SeM-specific IgGb (IgG4 and IgG7) response to infection but also significant increases in IgGa (IgG1), IgG(T) (IgG3 and IgG5) and a small increase in IgGc (IgG6). In comparison, a poor response by IgG1 (IgGa) to immunisation may have contributed to the failure of an extract vaccination to protect horses from subsequent *S. equi* challenge (Sheoran *et al.*, 1997). However, *in vitro* bactericidal activity of serum Ig does not necessarily indicate *in vivo* protection (Sheoran *et al.*, 1997; Timoney & Eggers, 1985).

Although IgG is the principal antibody in serum, serum IgA also contributes to the *in vitro* opsonophagocytosis of *S. equi* (Sheoran *et al.*, 1997) and secretory IgA is the predominant antibody in nasal secretions of adult horses (Sheoran *et al.*, 2000). SeM, Se18.9 and IdeE-specific mucosal IgA have been detected following clinical strangles and show a similar pattern of response to infection as serum IgG, peaking at 3 to 6 weeks post exposure and gradually declining over 4 to 6 months (Sheoran *et al.*, 1997; Timoney *et al.*, 2008; Tiwari *et al.*, 2007). Mucosal IgA may therefore contribute to immunity against *S. equi* infection, although a failure to react with a number of *S. equi* proteins has been noted (Galan & Timoney, 1985a). Ponies that received intranasal (and intramuscular) vaccination with mixed protein antigens, which conferred some protection from *S. equi* infection, generated antigen-specific IgG but only low concentrations of antigen-specific IgA in NP wash samples (Flock *et al.*, 2006). Mucosal IgG antibodies may therefore play a role in immunity and the failure of an extract vaccine to provide protection against strangles could

in part be due to the lack of an adequate mucosal IgG or IgA response (Sheoran *et al.*, 1997).

Since the port of *S. equi* entry is the naso- and oropharyngeal tonsils it is conceivable that a first defence would be most effective at the mucosal surface. One group of researchers, with some success, have focused their vaccine approach to block (at least in part) the function of adhesins that might be important in early adhesion and invasion events (Waller *et al.*, 2007). Immune horses appear to rapidly clear intranasally inoculated *S. equi* and also show no boost in serum antibody concentrations to some antigens previously recognised from *S. equi* infection, which may suggest that convalescent protective immunity occurs early on in the infectious process (Timoney *et al.*, 2007).

Immunity to *S. equi* infection may prove to be a complex process with many contributing facets as both intramuscular and intranasal routes of vaccination have demonstrated efficacy (Jacobs *et al.*, 2000; Timoney, 1993; Walker & Timoney, 2002). *S. zooepidemicus* is a normal commensal resident of the equine tonsil and very closely related to *S. equi*, but does not stimulate immunity that is cross-protective against *S. equi* infection (Bazely, 1942). For this reason, protective immunogens of *S. equi* are presumed, at least in part to be specific to this subspecies and more recent research into multi-component subunit vaccines has focused on the use of recombinant protein mixes that contain at least some *S. equi* unique components (Timoney *et al.*, 2007; Waller *et al.*, 2007). Comparative genomic analysis between *S. equi* and *S. zooepidemicus* will highlight potential vaccine targets that are specific to the former and this approach could reduce the time and cost associated with future vaccine development.

1.4 Phylogenetics of *Streptococcus zooepidemicus* and *Streptococcus equi*

The classification of streptococci has come a long way since 1903 when streptococci were first classified as either β-haemolytic or non-β-haemolytic according to their phenotype on blood agar. In 1933, Lancefield differentiated the β-haemolytic streptococci using serological tests based on carbohydrate 'group' antigens (Lancefield, 1933). In 1937 Sherman proposed a scheme for placing the streptococci into four categories based on the Lancefield typing and fermentation/tolerance profiles (Sherman, 1937). These were the pyogenic (pus forming) group (including the β-haemolytic streptococci Lancefield groups A-G), the viridans (oral) group (now viridans streptococci), the lactic group (now lactococcus) and the enterococci. This division of the streptococci is not appreciably different from that of today's identification systems except that now only the 2 first groups remain in the genus *Streptococcus*; the lactococcus and enterococci have been reclassified as separate genera.

The β-haemolytic streptococci have been separated into species and subspecies using a variety of techniques including DNA-DNA reassociation, 16s rRNA gene sequencing, whole-cell protein analysis, multilocus enzyme electrophoresis (MLEE), and phenotypic characteristics (Facklam, 2002). With the progression of DNA sequencing, multilocus sequence typing (MLST) has become established as a powerful molecular typing methodology to discriminate streptococcal isolates.

The pyogenic group of streptococci include several species of significant medical and veterinary importance. *Streptococcus agalactiae* (group B *Streptococcus* (GBS)) is the most common cause of neonatal sepsis (Facklam, 2002). *Streptococcus pyogenes* (group A *Streptococcus* (GAS)) causes bacterial pharyngitis, impetigo and a host of other infections

in man including severe invasive diseases. *Streptococcus canis* has the Lancefield group G antigen and, as the name suggests, is isolated frequently from dogs (Facklam, 2002). *S. equi* and *S. zooepidemicus* belong to the Lancefield group C β-hemolytic group of pyogenic streptococci and, based on 16s rRNA sequence, are most closely related to *S. pyogenes* and *S. canis* (Figure 1.1).

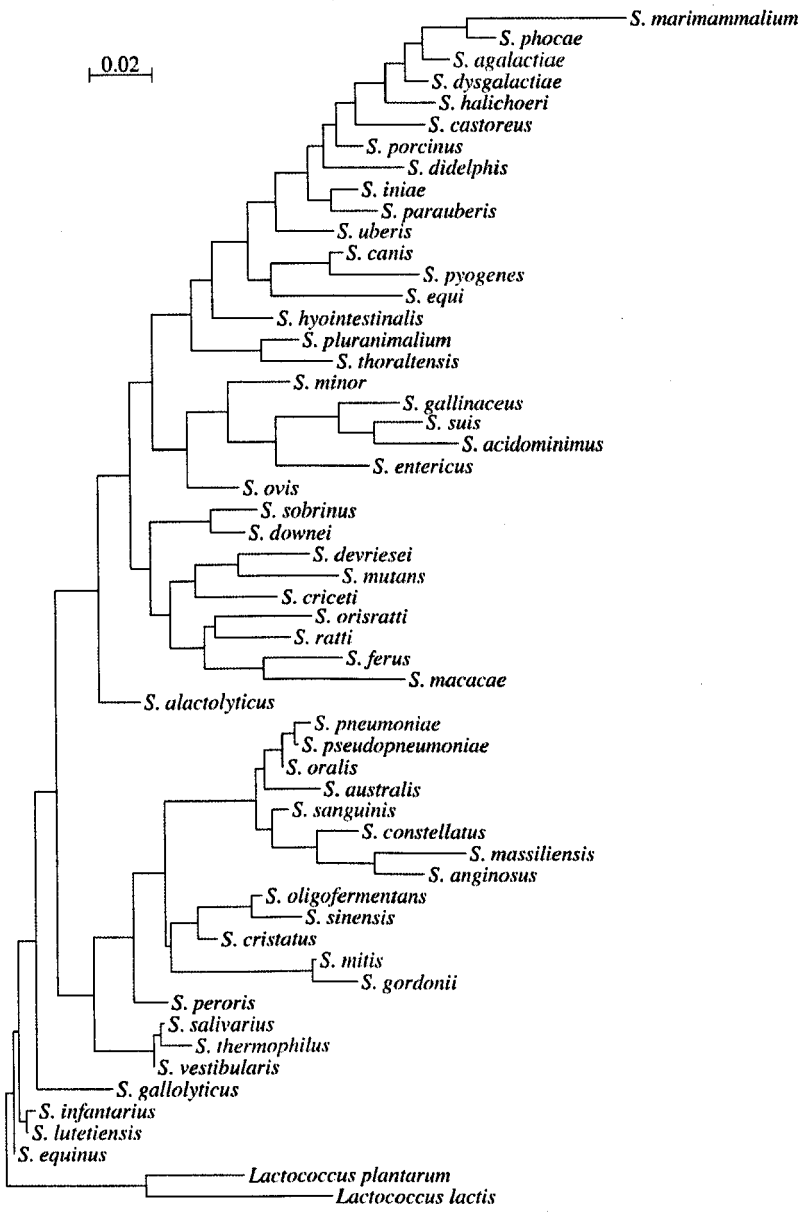


Figure 1.1 A maximum likelihood tree based on 16s rRNA showing the diversity of the streptococci

Figure reproduced with kind permission from Matt Holden, Sanger Institute. The *S. equi* subspecies share the same branch of the tree.

1.4.1 *Streptococcus equi* subspecies

S. equi and *S. zooepidemicus* are officially designated *S. equi* subspecies *equi* and *S. equi* subspecies *zooepidemicus* because *S. equi* was regarded as the archetype for *S. zooepidemicus*, based on DNA hybridisation studies (Farrow & Collins, 1984). However, later examinations using MLEE (Jorm *et al.*, 1994) and sequence analysis of 16S and 23S rRNA intergenic spacers (Chanter *et al.*, 1997) suggest the reverse scenario: that in fact *S. zooepidemicus* is the ancestor of *S. equi*. MLST provides the most compelling evidence yet to confirm *S. equi* as a descendant of *S. zooepidemicus* (Webb *et al.*, 2008) and a reclassification of these two subspecies seems appropriate to accurately reflect their origins.

A novel subspecies *S. equi subspecies ruminatorum* was recently isolated from milk samples from small ruminants with mastitis (Fernandez *et al.*, 2004). Based on 16S rRNA gene sequencing, *S. equi* and *S. zooepidemicus* are the closest phylogenetic relatives to *S. ruminatorum* with 98% DNA sequence identity. *S. ruminatorum* is β-haemolytic and possesses the Lancefield group C antigen but it does not correspond to *S. equi* or *S. zooepidemicus* in terms of its rRNA gene restriction patterns or biochemical reactions (Fernandez *et al.*, 2004). Unlike *S. equi* and *S. zooepidemicus, S. ruminatorum* is able to perform synergistic haemolysis with *Staphylococcus aureus* (Christie-Atkins-Munch-Petersen (CAMP) reaction) and hydrolyses hippurate but fails to acidify sucrose or methyl β-D-glucopyranoside. *S. ruminatorum* does however ferment lactose, sorbitol and ribose, a feature of *S. zooepidemicus* in > 90% , > 90% and 80-89% of strains respectively, but not a feature of *S. equi* (Holt, 1994). Like *S. equi* and *S. zooepidemicus*, *S. ruminatorum* is unable to ferment trehalose.

1.4.2 Population structure of *Streptococcus equi* and *Streptococcus zooepidemicus*

MLEE and MLST have highlighted the diverse population structure of *S. zooepidemicus* and the near clonal nature of the *S. equi* population. Using 247 Australian isolates cultured from horses (70 *S. equi* and 177 *S. zooepidemicus*), 3 clusters of 40 electrophoretic types (ET) of *S. zooepidemicus* and a unique ET-12 typical of *S. equi* were identified (Jorm *et al.*, 1994). Webb *et al* (Webb *et al.*, 2008) identified 130 unique sequence types (STs) from 277 isolates of varied geographical and temporal origin (Figure 1.2). All *S. equi* isolates were ST-179 or its single locus variant ST-151. An online MLST database is maintained at the University of Oxford (http://pubmlst.org/szooepidemicus), which currently holds 220 STs for *S. zooepidemicus* (n = 462) and 2 STs for *S. equi* (ST-179, n = 25, ST-151, n = 2, accessed 02.04.09) (Jolley *et al.*, 2004).

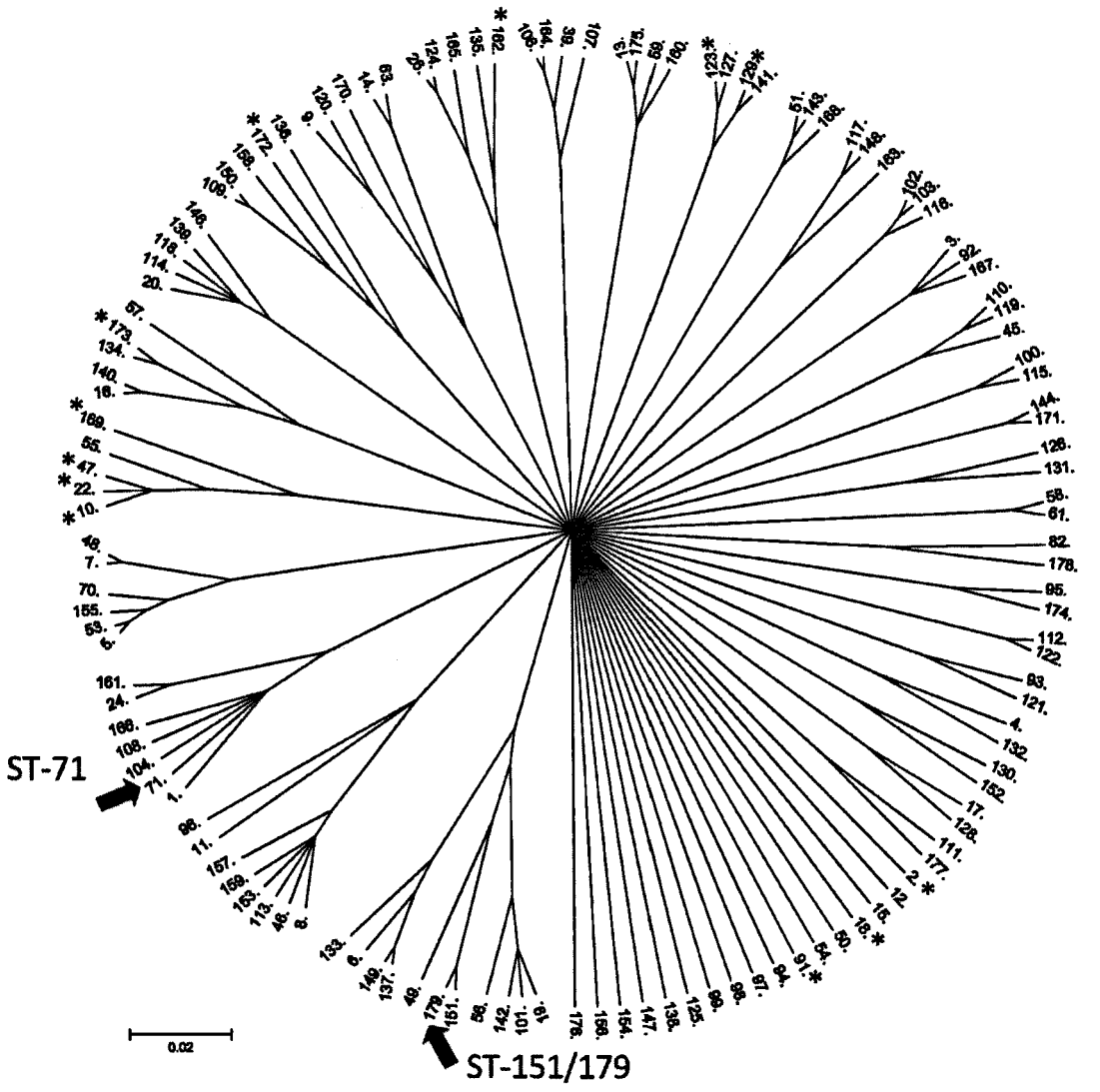


Figure 1.2 A majority-rules consensus tree

Tree generated from 12 independent runs of ClonalFrame, each with 250 000 iterations, and imported into MEGA to display as a radial tree (Webb *et al.*, 2008). The ST-151/179 cluster represents all *S. equi* isolates, including *Se*4047 (ST-179). The ST-71 cluster represents mostly *S. zooepidemicus* respiratory isolates (87%, n = 46), including *Sz*H70 (ST-1). Each ST in the ST-151/179 and ST-71 clusters is a single locus variant of the primary founder ST-179 and ST-71, respectively. *Isolates from outbreaks of acute fatal haemorrhagic pneumonia in dogs (Webb *et al.*, 2008). Figure reproduced with kind permission from Keith Jolley, Oxford University.

A propensity for recombination appears to be a feature of the *S. zooepidemicus* population and a contributing factor to the genetic diversity of this subspecies (Webb *et al.*, 2008). The eBURST diagrammatic representation of the *S. zooepidemicus* population based on MLST analysis (Figure 1.3) (Webb *et al.*, 2008) resembles similar population snapshots

generated for highly recombining species such as *Helicobacter pylori* (Turner *et al.*, 2007). Statistical tests of congruence show that the topologies of maximum-likelihood trees constructed using data from individual MLST loci are significantly different (Webb *et al.*, 2008) contrary to the phylogenetic congruence that occurs under entirely clonal evolution.

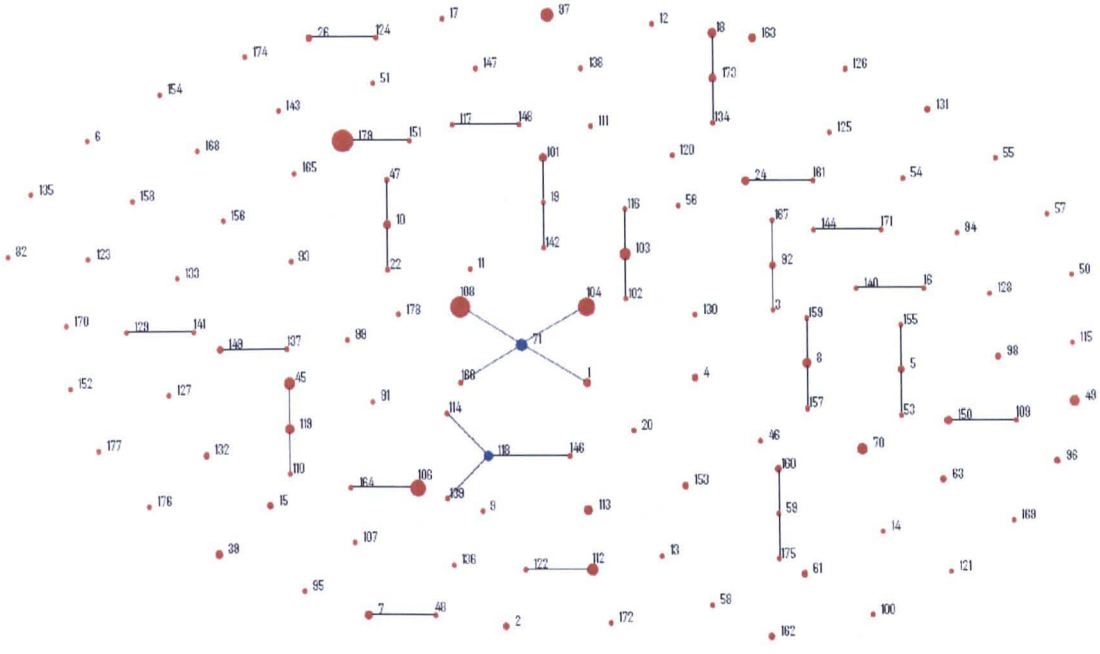


Figure 1.3 Population snapshot of the *S. zooepidemicus* group

Clusters of related STs and individual unlinked STs within the study sample are displayed as a single eBURST diagram (Webb *et al.*, 2008). Clusters of linked isolates correspond to clonal complexes. Primary founders (blue) are positioned centrally in each cluster, and subgroup founders are shown in red.

MLEE showed *S. equi* to be ET-12 (69 of 70 isolates), closely related to the *S. zooepidemicus* ET-13 and belonging to the electromorph cluster B with types 4 to 22 comprising 109 of 177 *S. zooepidemicus* isolates (Jorm *et al.*, 1994). MLST analysis revealed that some related *S. zooepidemicus* strains associated with particular disease types and surprisingly, isolates of *S. equi* shared a recent evolutionary ancestor with *S. zooepidemicus* isolates that were significantly associated with cases of uterine infection or abortion in horses ($P < 0.001$, Fischer's exact test, $n = 145$). *S. equi* clustered with 10

independent *S. zooepidemicus* isolates, 8 (80%) of which were associated with uterine infection and abortion. This cluster accounted for 18% (8 of 44) of all independent isolates from uterine infection or abortion in the MLST study compared with only 1% (2 of 203) of other independent *S. zooepidemicus* isolates studied (Figure 1.2) (Webb *et al.*, 2008).

S. zooepidemicus respiratory isolates formed 5 distinct phylogenetic clusters and included 40 of 46 isolates in the ST-71 complex, 13 of 17 isolates in the ST-8 cluster, 16 of 17 isolates in the ST-106/164 cluster, all 8 isolates in the ST-103 complex, all 11 isolates in the ST-119 complex and all 5 isolates in the ST-117/148 cluster (Figure 1.2 (Webb *et al.*, 2008)). Of a total of 145 *S. zooepidemicus* isolates associated with disease, a significantly higher proportion of isolates in the ST-71 cluster were recovered from cases of respiratory infection (rather than from other disease types) compared with those isolates that did not cluster in the ST-71 complex ($P < 0.001$, Fischer's exact test) (Webb *et al.*, 2008). A lack of association with respiratory colonisers does provide a possible explanation for the lack of evidence for colonisation of the equine respiratory tract by *S. equi* prior to invasion (Sweeney *et al.*, 2005). The sequenced *S. zooepidemicus* genome *Sz*H70 is ST-1 and clusters in the ST-71 respiratory complex, which contains independent isolates from the UK and USA collected over a long time period. This strain was chosen for genome sequence analysis before MLST data were available and is more distantly related to *S. equi* compared with other types. *Sz*H70 was chosen as a prevalent rRNA intergenic spacer/M-protein PCR-type A1HV1 (Newton *et al.*, 2008) isolated from a nasal swab taken from a healthy racehorse in Newmarket, England, in 2000 (Webb *et al.*, 2008). *Se*4047, ST-179, was isolated from a horse with strangles in the New Forest, England, in 1990 (Kelly *et al.*, 2006).

S. zooepidemicus STs associated with disease in dogs also often clustered together (Figure 1.2), including isolates from 3 UK outbreaks of acute fatal haemorrhagic pneumonia and

from 3 outbreaks of this disease in the USA. Some strain types isolated from dogs were also able to infect horses (Webb *et al.*, 2008).

1.4.3 Differentiation of *Streptococcus equi* strains

The absence of variation in housekeeping genes has required alternative methods to distinguish strains of *S. equi*. In the past, *S. equi* isolates were differentiated by PCR using enterobacterial repetitive intergenic consensus primers (Al-Ghamdi *et al.*, 2000). More recently sequence variation of the SeM gene, most likely driven by immune selection from the host, has allowed the discrimination of *S. equi* isolates (Anzai *et al.*, 2005; Kelly *et al.*, 2006). SeM is a virulence determinant involved in evasion of phagocytosis (Boschwitz & Timoney, 1994a; Galan & Timoney, 1987; Meehan *et al.*, 1998; Timoney *et al.*, 1997b). The short term variability of virulence genes or hypervariable genes has been used by others to enhance the discrimination of bacteria that lack variation in constitutive genes (van Loo *et al.*, 2002).

An online database (http://pubmlst.org/szooepidemicus/seM/) holds information on a total of 64 *S. equi* SeM alleles (accessed 02.04.09) including 43 reassigned in 2007 from two studies conducted in 2005 and 2006 (Anzai *et al.*, 2005; Kelly *et al.*, 2006; Waller & Jolley, 2007). Figure 1.4 shows the split decomposition analysis of the amino acid sequences of these 43 SeM alleles (Waller & Jolley, 2007). Both studies showed that groups of related isolates tended to be recovered from particular geographical locations, with allelic-types characteristic of North America and European outbreaks apparent. The SeM alleles of isolates from Brazil (alleles 40 and 41) and Australia (allele 15) were similar to those from European isolates, whereas SeM alleles of the majority of isolates from Japan (alleles 18, 19, 20 and 21) were related to alleles present in strains from the USA (alleles 24 and 25) (Waller & Jolley, 2007). This matched reports suggesting that an

outbreak of strangles in Tokachi, Japan originated from the transport of horses from Indiana, USA (Anzai *et al.*, 2005). SeM allele 9 is the most frequently identified allele type. The sequenced *S. equi* genome *Se*4047 has SeM allele type 3 which has been identified in 4 UK outbreaks and clusters with the European types (Kelly *et al.*, 2006).

SeM gene subtyping has been used to determine the source of new outbreaks, to follow changes in SeM sequence during the course of an outbreak and to differentiate between a vaccine strain and field strains of *S. equi* responsible for concurrent disease (Andrew Waller, unpublished data, (Kelly *et al.*, 2006)).

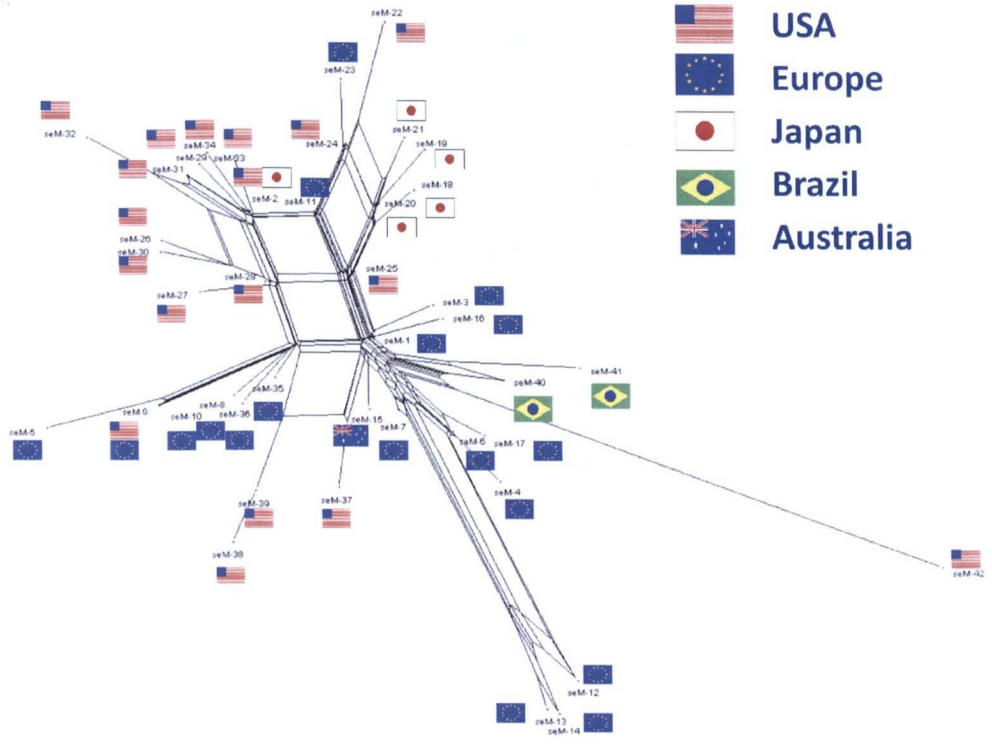


Figure 1.4 Split decomposition analysis of the amino acid sequences of 43 SeM alleles

Figure reproduced with kind permission from Andrew Waller, AHT and Keith Jolley, Oxford University (Waller & Jolley, 2007).

Chapter 2 Comparative and functional genomic analysis of *Streptococcus equi* and *Streptococcus zooepidemicus*

2.1 Introduction

2.1.1 Comparative genomic analysis

Comparative genomic analysis refers to the sequencing and comparison of multiple genomes of closely related organisms in order to identify differences in their gene content that may indicate the events that shaped their evolution. This technology has been applied to a wide genera of bacteria, including several human pathogens and their ancestral relatives. Common themes underpining the evolution of bacterial pathogens include gene gain, gene loss, gene change and often a combination of all three (Pallen & Wren, 2007). For example, *Yersinia pestis* (aetiological agent of plague) and *Shigella flexneri* demonstrate all three aspects of molecular divergence in their adaptation to new highly specialised niches, either as a successful blood-borne parasite of fleas and rodents (and occasionally humans) or as an intracellular pathogen of human intestinal epithelial cells, respectively (Parkhill *et al.*, 2001b; Wei *et al.*, 2003). *Y. pestis* has gained virulence factors that aid colonisation and persistence in the flea gut that have clearly contributed to its shift in host and transmission pathway, whilst *S. flexneri* has gained coding sequences (CDSs) on a large virulence plasmid, which have conferred the ability to invade epithelial cells (Sansonetti *et al.*, 1982; Wei *et al.*, 2003). Many other differences distinguish the genomes of *Y. pestis* and *S. flexneri* from their ancestors (*Yersinia pseudotuberculosis* and *E. coli*, respectively) including genome rearrangements and gene decay. For example, loss of gene function has contributed to a loss in the motility of both species (Parkhill *et al.*, 2001b; Wei *et al.*, 2003).

2.1.1.1 Gene gain

Horizontal gene transfer (HGT) is a potent source of gene gain and genomic variability between closely related bacteria. The extent of HGT was a revelation to the scientific community; however bacteria, inhabitants of the planet for over three and half billion years, have evolved a number of mechanisms to facilitate the uptake of foreign DNA in order to expand their potential for genetic adaptation. Genes can be acquired directly from other bacteria through contact-mediated transfer of DNA (conjugation), indirectly from bacteriophage vectors (transduction) or via the uptake of naked DNA from the surroundings (transformation). The ability to acquire DNA through these three mechanisms depends on the genetic make-up of the individual and in itself may be an adaptation of a particular species.

The genome of *Clostridium difficile*, for instance, contains a large proportion of integrative and conjugative elements (ICEs). As the name suggests, these mobile elements were putatively acquired by conjugation and encode an extensive array of genes likely to be involved in antimicrobial resistance as well as virulence, host interaction and the production of surface structures (Sebaihia *et al.*, 2006). Virulent, multidrug-resistant strains of *C. difficile* have been responsible for recent outbreaks of antibiotic-associated diarrhoea. Related ICEs often share the same integration sites, and thereby increase the variability at such sites within both bacterial species and genera (Burrus & Waldor, 2004). Often the genes encoding the apparatus enabling conjugation are located on plasmids.

Transduction has played an important role in the emergence of highly virulent clones of GAS and many other pathogens (Brussow *et al.*, 2004; Canchaya *et al.*, 2003). GAS infect humans only and are a common cause of pharyngitis ('strep throat'), but more severe forms of group A streptococcal disease (e.g. necrotizing fasciitis or 'flesh-eating disease' and streptococcal toxic shock syndrome) re-emerged in the 1980s. Comparative genomic

analyses revealed a large array of prophage in the genomes of the more virulent M-types of GAS associated with these outbreaks of severe invasive disease (Aziz *et al.*, 2005; Beres *et al.*, 2002; Ikebe *et al.*, 2002; Nakagawa *et al.*, 2003). The prophage elements carry genes that are superfluous for bacteriophage replication within distinct 'cargo compartments' at one end of the prophage genome, which encode DNAses, streptococcal phospholipase A_2 (Sla) and streptococcal pyrogenic exotoxins. The latter are superantigen-like extracellular proteins, which elicit a massive non-specific immunological response in the host and contribute to pathogenesis.

Bacteriophage have also contributed genetic diversity to pathogens as distinct as *E. coli* (shiga toxin and type-III secretion effectors), *Corynebacterium diphtheriae* (diphtheria toxin), *Vibrio cholerae* (cholera toxin) and *S. aureus* (staphylococcal enterotoxin A, exfoliatin A and Panton-Valentine leukocidin) (Pallen & Wren, 2007). Interestingly, recombination between phage modules appears common (Brussow *et al.*, 2004; Canchaya *et al.*, 2003) and so the potential for mobilization of cargo carried by phage is increased, which adds an extra dynamic to the fluidity of these phage-encoded genes among bacterial pathogens.

Regions of atypical nucleotide content indicative of HGT are referred to as genomic islands where clusters of contiguous genes are present in some related strains or species but not in others. These may encode integrases (enzymes that integrate the island into the host DNA) and are often associated with transfer RNA loci. Pathogenicity islands (PIs) encode proteins that contribute a virulence phenotype (Pallen & Wren, 2007). However, the mechanism of mobilization of these genomic islands is not always apparent. The PIs of *S. aureus* are thought to have arisen following a misrecombination event that could have led to the replacement of a segment of phage DNA with a chromosomal segment (Feng *et al.*,

2008; Yarwood *et al.*, 2002). These loci require 'helper' phage for their mobilisation and can be transduced at a very high frequency (Lindsay *et al.*, 1998; Ruzin *et al.*, 2001).

In *Salmonella*, the causative agent of food-borne gastroenteritis and typhoid fever, the presence of virulence genes on numerous PIs has conferred the ability of these organisms to invade and survive in eukaryotic cells, and to cause disease (Whittam & Bumbaugh, 2002). *E. coli* O157:H7, implicated in major food-borne outbreaks of disease, has over 1.4 Mb of specific DNA relative to the laboratory commensal strain *E. coli* K-12. This O157:H7 specific DNA is organised into approximately 180 discrete regions (O-islands) with the potential to encode over 1500 proteins, approximately 10% of which are assumed to have virulence-related functions (Hayashi *et al.*, 2001; Perna *et al.*, 2001; Whittam & Bumbaugh, 2002).

Plasmids and chromosomal elements have shaped the evolution of *Y. pestis*. Like its close relatives *Y. pseudotuberculosis* and *Yersinia enterocolitica*, *Y. pestis* is host to a 70 kb virulence plasmid pYV/pCDI. In addition, it also carries two other plasmids, pPst/pPCP1 and pFra/pMT1 which are not present in the other *Yersinia* species. Together, these plasmids, and a high pathogenicity island (HPI), encode virulence factors involved in bacterial adhesion, injection of proteins into the host cell (type-III secretion molecules), invasion (plasminogen activator Pla), and the production of a siderophore yersiniabactin, important for iron acquisition (Bearden *et al.*, 1997; Cowan *et al.*, 2000; Parkhill *et al.*, 2001b).

Evidence for gene duplication, another source of gene gain, has been identified in the genomes of *S. equi* and *S. zooepidemicus*, which contain 2 near-identical paralogues of *hasC* and *glmU*, genes encoding enzymes involved in the biosynthesis of hyaluronate (HA) capsule prescursors (Blank *et al.*, 2008). This duplication event is likely to improve HA

productivity by gene dosage. Similar gene expansion events are likely to have contributed to the evolution of *S. equi* (Karlstrom *et al.*, 2006) and are evident in many bacterial species encoding a diverse set of functions (Hollingshead *et al.*, 1994; Tettelin *et al.*, 2001). Half of the genes (52%) present in *Mycobacterium tuberculosis* arose from gene-duplication events leading to extensive functional redundancy (Tekaia *et al.*, 1999).

Expansion of insertion sequence (IS) elements is a defining feature of organisms that have encountered an 'evolutionary bottleneck'. This is common to many recently emerged pathogens that have entered a new niche with a decreased population size. In any population, ISs transpose through the genome to novel sites, but many of the resultant insertions disrupt CDSs or regulatory components and may confer a selective disadvantage. In normal circumstances, bacteria carrying them are often outcompeted by others in the population. However, for those that have encountered an 'evolutionary bottleneck' the selective disadvantage may not exist in the new environment and/or the mutation may be less likely to be competed out or repaired through recombination due to a small population size. The result is a build up of ISs in the population. Accumulation of ISs is particularly evident in *Y. pestis* (140 ISs), *B. pertussis* (261 ISs), *B. mallei* (171 ISs) and Shigella flexneri (314 ISs) (Jin *et al.*, 2002; Nierman *et al.*, 2004; Parkhill *et al.*, 2001b; Parkhill *et al.*, 2003).

2.1.1.2 Gene loss

Despite the widespread occurrence of HGT, bacterial genomes tend to remain about the same size. This is because gene gain is usually accompanied by gene loss as genes unnecessary for survival in a new niche are no longer maintained by natural selection. This phenomenon of genome reduction is well demonstrated by the recently emerged pathogens that have changed lifestyle and often moved from a more diverse existence to

occupy a more restricted host-associated niche. For example, the genomes of *Salmonella enterica* serovar *typhi*, *Y. pestis*, *Bordetella pertussis*, *Burkholderia mallei* and *Mycobacterium leprae* show substantial genome decay when compared to their more flexible ancestors (Cole *et al.*, 2001; Maurelli, 2007; Nierman *et al.*, 2004; Parkhill *et al.*, 2001a; Parkhill *et al.*, 2001b; Parkhill *et al.*, 2003); the latter 3 species have evolved almost solely through gene loss. In most cases, the loss of gene function appears to have occurred relatively recently as small mutations have introduced frame shifts or early stop codons (pseudogenes). With time, larger regions of the gene or genome are lost. An abundance of ISs contributes to gene loss following recombination between these perfect repeat elements. *M. leprae* demonstrates extreme genome downsizing relative to other mycobacteria with less than half of the genome encoding functional protein-coding sequences. Gene deletion and decay have eliminated many important metabolic activities including siderophore production, part of the oxidative and most of the microaerophilic and anaerobic respiratory chains, and numerous catabolic systems and their regulatory circuits (Cole *et al.*, 2001). Similar severe decay is seen in other obligate intracellular parasites (Andersson *et al.*, 1998), which suggests that these organisms encode 'just enough' activity to permit intracellular growth.

Situations are likely to arise where gene decay results from positive selection. This concept falls under the umbrella of 'pathoadaptation' (Maurelli, 2007). A newly minted pathogen, freshly equipped with virulence factors may require further modifications in order to better survive a new niche. Genes that are no longer compatible with the novel lifestyle of the pathogen are selectively altered or inactivated. These have been referred to as 'antivirulence genes' when the gene in question affects the pathogenicity of the organism (Maurelli, 2007). As an example, strains of shigella lack ompT; a gene present in closely related non-pathogenic *E. coli*. Introduction of ompT, which encodes a surface

protease, suppresses shigella virulence by disrupting intercellular spread (Nakata *et al.*, 1993).

2.1.1.3 Gene change

Gene change refers to modifications that affect the sequences or order of the existing genes. A common feature of IS accumulation, in addition to gene decay, is genome rearrangement and translocation (Turlan & Chandler, 1995). *Bordetella pertussis, Y. pestis, S. flexneri 5b* and *Burkholderia mallei* all demonstrate varying amounts of chromosomal rearrangement mediated by ISs. *B. pertussis*, in comparison to *Bordetella bronchoseptica*, has nearly 150 individual rearrangements, unprecedented for congeneric bacteria. Recombinations between ISs in *Y. pestis* are reflected by striking anomalies in GC bias and also occur during the growth of the organism *in vitro*. The effects of these frequent intra-genomic recombination events is a fluid genome but their influence on the biology of the organism is unknown (Parkhill *et al.*, 2001b).

Adaptation to a new niche is a continuous process. Newly acquired virulence traits, along with the pre-existing core genes, continue to undergo selection. Mutations within individual genes must also contribute to the initial development of gene functions as well as their subsequent fine tuning. Nonsynonymous changes and natural selection of the encoded proteins most likely determines the fate of new alleles.

High levels of polymorphism have been observed at loci involved with host-pathogen interactions in many different bacterial species (Marciel *et al.*, 1997; McShan *et al.*, 2008; Nassif *et al.*, 1993; Perna *et al.*, 1998; Salama *et al.*, 2000; Theisen *et al.*, 1995; Zhang & Wise, 1996). Genetic variation often occurs in surface adhesins. Uropathogenic *E. coli* express a variant form of FimH, a fimbrial adhesin, which confers an advantage for colonisation of the bladder through increased binding to mannosylated glycoproteins

(Sokurenko *et al.*, 1998). *S. zooepidemicus* demonstrates hyper-variability in the sequence of a protective adhesin (Fan *et al.*, 2008; Timoney *et al.*, 1995; Timoney *et al.*, 1997b). Where organisms such as *S. zooepidemicus* and *S. equi* have evolved a transmission pattern that relies on persistent infection, mechanisms for evading the host adaptive immunity are particularly important. The ratio of nonsynonymous to synonymous nucleotide changes in the *seM* gene of *S. equi* is extremely high suggesting that strong immune selection is directed against epitopes in the N-terminus of the antiphagocytic SeM protein (Kelly *et al.*, 2006). Moreover, sequence changes occurred over time (Kelly *et al.*, 2006) and in some cases truncations arose in this region of *seM* in *S. equi* isolates that remained in the GPs of asymptomatic carriers for prolonged periods of time (Chanter *et al.*, 2000).

Sequence changes can be introduced by random mutation; however recombination between closely related sequences in closely related strains may also contribute. Recombination is a predominant feature of *S. zooepidemicus* (Webb *et al.*, 2008), *Helicobacter pylori* (Aras *et al.*, 2003) and other species such as *Neisseria meningitidis*, *Haemophilus influenzae* and *Streptococcus pneumoniae* that are naturally competent and display hyper-variable chromosomal regions (Alm *et al.*, 1999; Nassif *et al.*, 1993; Read *et al.*, 2000). Not all variation is expected to be adaptive. Some may present neutral variation between strains. Still others may be deleterious but either have not yet been eliminated by selection or cannot be eliminated because of linkage constraints.

Many mechanisms of antigen and phase variation exist that allow bacteria to undergo frequent and usually reversible phenotypic changes resulting from genetic or, more rarely, epigenetic alterations at specific genetic loci. *Borrelia burgdorferi*, the causal agent of Lyme disease, escapes the humoral immune system by generating variants expressing different antigenic forms of the VlsE surface lipoprotein. A functional copy of *vlsE* can recombine with any one of 15 silent cassettes to create new versions of the lipoprotein

(gene conversion) (Bankhead & Chaconas, 2007; Wisniewski-Dye & Vial, 2008; Zhang & Norris, 1998).

A mechanism of phase variation displayed by *S. pyogenes* switches the production of the adhesin SclB on and off through a process of slipped-strand misrepairing in pentanucleotide coding repeats (Rasmussen & Bjorck, 2001), a process that throws the CDS in and out of frame. Site-specific inversion of an element that usually contains a promoter and is flanked by short inverted repeat sequences can be catalysed by recombinases. The inversion event acts as a switch triggering the expression of a gene initially silent or impeding the expression of a gene initially expressed (Wisniewski-Dye & Vial, 2008). *Bacteroides fragilis* uses DNA inversion to modulate more than 20 genetic loci, which contain genes that encode bacterial surface proteins, polysaccharides and components of regulatory systems (Cerdeno-Tarraga *et al.*, 2005). Phase variation can have more far reaching effects by impacting on global regulation (van der Woude & Baumler, 2004). Coordinate phase variation of key *S. pyogenes* surface proteins is achieved through change in the central regulator, *mga*, required for expression of M protein and C5a peptidase (Bormann & Cleary, 1997).

2.1.2 Genetic variation and virulence: *Streptococcus equi* and *Streptococcus zooepidemicus*

This introduction describes the pre-existing knowledge of genetic determinants and putative virulence factors that are shared between *S. equi* and *S. zooepidemicus* or distinguish *S. equi* from its closely related proposed ancestor (Table 2.1). To relate these molecular factors to the processes of infection these are described under 3 headings, i) adhesion, ii) immune evasion and iii) tissue damage, dissemination and nutrient acquisition. However, many presented attributes offer functions covering more than one category and have been explained accordingly.

Table 2.1 Genetic determinants and potential virulence factors of *S. equi* and *S. zooepidemicus*

Protein/factor				
***S. equi* (*Se*)**	***S. zooepidemicus* (*Sz*)**	**Function**	**Features**	**Protection against *S. equi* infection following protein vaccination**
Fne	Fnz	CL- and Fn-binding	Surface protein in *Sz*, secreted, truncated protein in *Se*	Partial protection in mice when combined with Sfs and Eag
FneB	Fnz2	CL- and Fn-binding	Surface protein	No protection in mice
Sfs	Sfs (detected in 85% strains only, n = 48)	Fn- binding	Secreted protein	Partial protection in mice when combined with Fnz and Eag
Cne	Cne	CL- binding	Surface protein	Partial protection in mice enhanced when used in combination with Eag, partial protection in horses in combination with Eag and SclC
FneC	?	CL- binding	Surface protein	Not tested
FneD	?	CL- binding	secreted, truncated protein in *Se*	Not tested
FneE	?	CL- and Fn-binding	Surface protein	Not tested
FneF	?	CL- binding	Surface protein	Not tested
SclC	?	Unknown	CL-like surface protein	Partial protection in mice, partial protection in horses in combination with Eag and Cne
SclD	?	Unknown	CL-like surface protein	Not tested
SclE	?	Unknown	CL-like surface protein	Not tested
SclF	?	Unknown	CL-like surface protein	Not tested
SclG	?	Unknown	CL-like surface protein	Not tested
SclH	?	Unknown	CL-like surface protein	Not tested
SclI	?	Unknown	CL-like surface protein	Not tested
SeM	SzM	SeM: Antiphagocytic, IgG-, Fg-binding, adherence SzM: unknown	Surface protein	No protection in horses
SzPSe	Szp	SzPSe: Fg-binding SzP: adhesin	Surface protein	Not tested

Table 2.1 continued

Protein/factor		Function	Features	Protection against *S. equi* infection following protein vaccination
S. equi (*Se*)	*S. zooepidemicus* (*Sz*)			
Se18.9	-	Fg- and factor H-binding, antiphagocytic	Secreted protein that also associates with the cell surface	Not tested
Hyaluronate capsule		Antiphagocytic, adherence?	Capsular polysaccharide	Non-immunogenic
SeeH	-	Superantigen	Phage associated exotoxin: induces equine PBMC proliferation* and pyrexia in rabbits	Not tested
SeeI	-	Superantigen	Phage associated exotoxin: induces equine PBMC proliferation and pyrexia in rabbits/horses	Protection against the pyrogenic effects of SeeI but not tested for protection against *S. equi* infection
SeeL	SeeL (only detected in 1 strain out of 31 strains tested)	Superantigen	Phage associated exotoxin	Not tested
SeeM	SeeM (only detected in 1 strain out of 31 strains tested)	Superantigen	Phage associated exotoxin	Not tested
Eag	Zag	IgG-, α_2M-, albumin-binding	Surface protein	Partial protection in mice enhanced when used in combination with Fnz and Sfs, also partial protection in mice when used in combination with Cne, partial protection in horses in combination with Cne and SclC
IdeE	IdeZ	Antiphagocytic IgG endopeptidase (not active against equine IgG4)	Secreted protein	Not tested
SLS	SLS	β-haemolysis	Secreted streptolysin S-like toxin	Not tested
Shp	?	Haem-binding and transfer	Surface protein	Not tested
HtsABC	?	Haem-binding and uptake transfer	ABC transporter	Not tested
MtsABC	MtsABC	Putative Mn/Fe binding and uptake	ABC transporter	Not tested

Table 2.1 continued

Protein/factor				
***S. equi* (*Se*)**	***S. zooepidemicus* (*Sz*)**	**Function**	**Features**	**Protection against *S. equi* infection following protein vaccination**
Hpr	Hpr	Putative regulation of sugar uptake	Histidine containing phosphocarrier protein of the streptococcal phosphotransferase system	Not tested
LppC	LppC	Acid phosphatase	Lipoprotein	Not tested
Pst	Pst	Putative regulation of sugar uptake	Phosphoenolpyruvate -protein phosphotransferase	Not tested
HAP	HAP	Putative ecto threonine kinase, putative regulatory role in hyaluronate capsule shedding through auto-phosphorylation	Hyaluronate associated lipoprotein	Partial protection in mice but no protection in horses

CL: collagen; Fn: fibronectin; Fg: fibrinogen; α_2M: α_2-macroglobulin. * Not shown to induce proliferation of equine PBMCs in this study.

2.1.2.1 Adhesion

Attachment to host cell surfaces is an essential first step in bacterial infection, which in Gram-positive bacteria is largely mediated by cell wall-anchored surface proteins that bind to components of the ECM such as fibronectin, collagen and fibrinogen (Patti *et al.*, 1994). *S. equi* and *S. zooepidemicus* are likely to share factors that contribute to this early adhesive event. However, whilst *S. zooepidemicus* can remain at the respiratory surface for some time as evidenced by the regular isolation of this subspecies from the equine nasopharynx, *S. equi* rapidly invades the epithelium and subepithelial follicles before being translocated within a few hours to the efferent LNs. Differences therefore exist in the invasive capabilities of these 2 subspecies and possibly their competence for persistent colonisation, at least of the mucosal surface in the oro- and nasopharynx. Fibronectin-

binding proteins of *S. pyogenes* and *S. aureus* contribute to integrin-mediated cellular invasion (Hauck & Ohlsen, 2006; Schwarz-Linek *et al.*, 2004) but such events have yet to be proven in *S. equi*. Fne produced by *S. equi* has been shown to re-shape host cells through an interaction with host integrins and ECM components (see below and section on immune evasion). *S. equi* may therefore use an as yet undetermined active process to mediate the early invasion event (possibly through factors described here) and/or it may rely on an indirect process of dissemination (see tissue damage, dissemination and nutrient acquisition below).

Both *S. equi* and *S. zooepidemicus* harbor at least 3 genes encoding fibronectin-binding proteins (Table 2.1). In subspecies *zooepidemicus*, 2 of these proteins, Fnz (Lindmark *et al.*, 1996) and Fnz2 (Hong, 2005), are cell wall bound while the 3rd, Sfs (Lindmark & Guss, 1999) has characteristics of a secreted protein. In subspecies *equi*, the corresponding proteins are denoted Fne (Lindmark *et al.*, 2001), FneB (Lannergard *et al.*, 2005) and Sfs. However, a frameshift mutation has occurred in the *fne* gene during the evolution of *S. equi* and as a result Fne is truncated relative to Fnz. Consequently, Fne does not have a cell wall anchor and is secreted from the cell but also lacks the 2nd fibronectin-binding domain of Fnz, which reduces the opportunity for fibronectin binding by *S. equi* via 2 potential mechanisms. Firstly, by a reduced surface bound fibronectin binding capacity and secondly, because Fne is released into the extracellular environment fewer free fibronectin targets may be available for binding. Sfs may act in a similar fashion to reduce the availability of fibronectin ligands and although *sfs* is a constant feature of *S. equi* (n = 50), 15% of *S. zooepidemicus* strains (n = 48) lack this gene (Lindmark & Guss, 1999). *S. equi* strains typically bind significantly lower quantities of fibronectin than those of *S. zooepidemicus* (Lindmark *et al.*, 2001). A recombinant form of another predicted cell wall-anchored protein FneE identified in *S. equi* also binds fibronectin (Table 2.1) (Lannergard, 2006), although fibronectin binding by *S. equi* appears to be mostly mediated

by the cell wall bound FneB (Lannergard *et al.*, 2005). However, mice vaccinated with FneB were not protected against *S. equi* infection (Flock *et al.*, 2006), although vaccination with Fnz and Sfs in combination with one other protein did confer some protection (Flock *et al.*, 2004). It is worth noting that reduced fibronectin-binding capacity of GAS has resulted in increased virulence through enhanced bacterial dissemination (Nyberg *et al.*, 2004) and the same may apply for *S. equi*.

Fnz/Fne and Fnz2/FneB also bind collagen, the most abundant structural component of ECM. Secretion of Fne may therefore also reduce the ability of *S. equi* relative to *S. zooepidemicus* to directly interact with host tissue via collagen binding. Another 5 *S. equi* putative cell wall-anchored proteins are involved in collagen binding: Cne, FneC, FneD, FneE and FneF (Table 2.1) (Lannergard *et al.*, 2003; Lannergard, 2006), although FneD, like Fne, is interrupted by an early stop codon and is likely to be secreted as a truncated product. Cne belongs to the Cna family of collagen-binding proteins (Nallapareddy *et al.*, 2003; Patti *et al.*, 1992; Rich *et al.*, 1999; Sato *et al.*, 2004; Xu *et al.*, 2004) and contributes to 35% of the collagen binding activity of *S. equi in vitro*. The *cne* gene was a feature of all strains of *S. equi* (n = 13) and also all strains of *S. zooepidemicus* (n = 13). Cne as a subunit vaccine alone or in combination with 2 other recombinant proteins conferred partial protection against *S. equi* infection in mice and horses respectively (Flock *et al.*, 2006; Waller *et al.*, 2007). FneC, FneD, FneE and FneF were identified as putative paralogues of Fne by BLAST analysis of the publicly accessible *Se*4047 genome and their presence in *Sz*H70 is addressed in the results section of this chapter.

A family of 7 collagen-like cell surface proteins (SclC – SclI) have also been identified in 2 strains of *S. equi*, *Se*4047 and 1866 (Karlstrom *et al.*, 2004; Karlstrom *et al.*, 2006) but their presence in *S. zooepidemicus* has not been previously examined (Table 2.1). A common feature of these proteins are GXY-triplet repeats (collagen-like) and variable N-

terminal regions (V regions) that show no similarity to known proteins. The biological role of these Scl proteins in *S. equi* is yet to be investigated but a collagen-like protein of *S. pyogenes* interacts with mammalian collagen-binding integrins (Humtsoe *et al.*, 2005) and similar host interactions may occur through this expanded family of Scls in *S. equi*. Sera from horses previously diagnosed with strangles have antibodies against these proteins suggesting their expression by *S. equi* during infection (Karlstrom *et al.*, 2006). In addition, SclC as a subunit vaccine alone or in combination with 2 other recombinant proteins conferred partial protection against *S. equi* infection in mice and horses, respectively (Flock *et al.*, 2006; Waller *et al.*, 2007).

Fibrinogen is an abundant plasma protein that increases in concentration during the acute phase of *S. equi* infection and can serve as a substrate for bacterial adhesins by coating trauma sites or as a component of the ECM. *S. equi* produces 2 cell wall-anchored proteins SeM (also known as FgBP) and SzPSe that bind strongly to equine fibrinogen (Table 2.1) (Meehan *et al.*, 2002; Timoney *et al.*, 1997b), although pre-treatment of *S. equi* with SeM-specific antibodies reduces fibrinogen binding by about 70% (Boschwitz & Timoney, 1994b) and *S. equi* mutants lacking SeM fail to bind fibrinogen (Meehan *et al.*, 2001), suggesting that SeM is more important for fibrinogen binding than SzPSe. The adherence of *S. equi* to cheek and tongue cells isolated from adult ponies could be inhibited by pre-incubation of cells with purified SeM and antiserum raised to SeM inhibited adherence to the same extent as antiserum raised against whole *S. equi* cells (Srivastava & Barnum, 1983).

Both SzPSe and SeM have homologues in *S. zooepidemicus* named Szp and SzM, respectively (Table 2.1). SzPSe is an allelic variant of Szp, a hypervariable protective typing antigen of *S. zooepidemicus* (Timoney *et al.*, 1995; Timoney *et al.*, 1997b) that functions as an adhesin and mediates the binding of *S. zooepidemicus* to HEp-2 cells (Fan

et al., 2008). SeM has a unique N-terminal domain relative to its allelic variant in *Sz*H70 (Kelly *et al.*, 2006). As this N-terminal region is necessary for the binding of SeM to fibrinogen (and IgG), the replacement or modification of this region may represent an important evolutionary adaptation by *S. equi* and may have contributed to its increased virulence over *S. zooepidemicus* (Galan & Timoney, 1987; Timoney *et al.*, 1997b). *S. equi* SeM mutants show decreased virulence in a mouse model (Meehan *et al.*, 2001) while the contribution of SzPSe to the virulence of *S. equi* has not been investigated. Se18.9 is a secreted fibrinogen-binding protein (and factor H-binding protein) that is unique to *S. equi* and absent from *S. zooepidemicus* (Table 2.1). However, this protein also associates with the cell surface through an unknown mechanism (Tiwari *et al.*, 2007).

The HA capsule of *S. zooepidemicus* and *S. equi* may influence adherence of these bacteria to host cells, although conflicting data showed that higher levels of HA increased binding to HeLa cells by *S. zooepidemicus* (Wibawan *et al.*, 1999) and decreased binding to tongue cells of ponies by *S. equi* (Srivastava & Barnum, 1983). An acapsular *S. equi* HA synthase deletion mutant (*ΔhasA*) did however show reduced clearance from *in vitro* air-interface respiratory epithelium cultures (Josh Slater, unpublished data). HA capsule may act as a direct ligand for host receptors (Harrington *et al.*, 2002; Wessels, 2000), mask the active sites of adhesins or exert an influence through its negative charge on the conformation and functionality of surface exposed proteins involved in adhesion (Timoney, 2004).

2.1.2.2 Immune evasion

Fever, neutrophilia and increased plasma fibrinogen are symptoms of acute phase strangles that may be triggered as a consequence of superantigen production, which is a feature of *S. equi* but not generally *S. zooepidemicus* (Anzai *et al.*, 1999a). By screening a *S. equi* chromosomal phage library for superantigen activity, 2 superantigens (SeeH and SeeI,

initially referred to as *sePE-H* and *sePE-I*) were identified in strain CF32 (Table 2.1) (Artiushin *et al.*, 2002). SeeH and SeeI are near identical orthologues of SpeH and SpeI, the pyrogenic superantigens of *S. pyogenes*. *seeH* and *seeI* were present in 7 out of 7 *S. equi* strains (Artiushin *et al.*, 2002) but absent from all *S. zooepidemicus* strains tested (n = 7). Later, 2 orthologues of the *S. pyogenes* superantigens SpeL and SpeM were identified by data mining of the *Se*4047 genome (Table 2.1) (Alber *et al.*, 2005) and proved to be a consistent feature of *S. equi* strains (n = 17). An earlier study failed to detect these CDSs in 8 strains of *S. equi* but did not classify the subspecies of the investigated *S. equi* isolates (Proft *et al.*, 2003). Genes encoding SeeL and SeeM were identified in one strain of *S. zooepidemicus* (out of 31 screened) isolated from bovine mastitis (Alber *et al.*, 2005) and another *S. zooepidemicus* isolate recovered from a human with toxic shock-like syndrome gave evidence of superantigen production although the isolate was not screened for *seeL* or *seeM* (Korman *et al.*, 2004). The location of *seeH* + *seeI* is adjacent to putative phage genes similar to those found in a defective phage sequence in *S. pyogenes* M1 and *seeL* + *seeM* are next to a putative phage encoded cell wall hydrolase, which points to their acquisition via phage mediated horizontal transfer (Alber *et al.*, 2005; Artiushin *et al.*, 2002). Greater than 98% amino acid sequence identity with the *S. pyogenes* counterparts suggests that an exchange of superantigens has occurred between these different species of streptococci.

Bacterial superantigens are small secreted toxins that simultaneously bind and cross-link major histocompatibility class (MHC) II antigens (outside of the normal antigen-binding interface) and T cell receptors leading to a non-specific stimulation of large numbers of T cells, a resultant pro-inflammatory cytokine storm and corruption of the innate and adaptive immune response (Fraser & Proft, 2008). This is likely to represent an important mechanism employed by *S. equi* during the early stages of infection to avoid host responses that might normally reduce the pathogen load. Superantigen production may

also enhance abscess development. SeeH and SeeI both induced strong mitogenic responses in horse peripheral blood mononuclear cells (PBMCs) and pyrexia in rabbits, however only SeeI was pyrogenic in horses. Horses immunised with SeeI or recently recovered from strangles had serum antibodies specific to this mitogen (and SeeH in convalescents) and were resistant to the pyrogenic effects of SeeI (Artiushin *et al.*, 2002).

Studies with *S. pyogenes* demonstrated stimulation of pro-inflammatory cytokines following interaction of M-protein with PMNs and monocytes (Pahlman *et al.*, 2006). SeM (see above) might prove to misdirect the immune response in a similar fashion. As the major cell wall-associated protein of *S. equi* cells grown *in vitro* (Meehan *et al.*, 1998), SeM contributes significantly to the resistance of *S. equi* to phagocytosis (Boschwitz & Timoney, 1994a; Boschwitz & Timoney, 1994b; Chanter, 1994; Meehan *et al.*, 2001; Timoney *et al.*, 1997b), a particular feature of this bacterium which is demonstrated by the accumulation of long chains of extracellular streptococci surrounded by large numbers of degenerating neutrophils in the abscesses of strangles. The antiphagocytic action of SeM results from the binding of fibrinogen and the Fc domain of IgG; the latter inhibits deposition of C3b on the bacterial surface. SeM binds specifically to IgG4 and IgG7, the most prevalent forms of IgG in serum and mucosal secretions (Sheoran *et al.*, 2000) to inhibit complement activation via the classical pathway (Lewis *et al.*, 2008a). Since IgG4 and IgG7 are likely to interact with Fc receptors on effector cells, SeM may also inhibit phagocytosis directly (Lewis *et al.*, 2008b). Fibrinogen-binding inhibits fixation of C3 on human streptococci (Chhatwal *et al.*, 1985), but a similar inhibitory effect was not demonstrated in *S. equi* (Boschwitz & Timoney, 1994b) and so the process used by fibrinogen binding to protect against phagocytosis is unclear. The promotion of bacterial aggregation by SeM offers another mechanism to avoid phagocytosis (Meehan *et al.*, 2001), a property critical for resistance to phagocytosis (and adherence) in *S. pyogenes* (Frick *et al.*, 2000).

Another IgG-binding protein has been identified in both *S. equi* (Eag) and *S. zooepidemicus* (Zag) that could contribute to immune evasion in a similar way to SeM (Table 2.1). Zag binds horse IgG with a much higher affinity than SeM and also binds IgG from a broader range of host species (Jonsson *et al.*, 1995). *S. equi* and some isolates of *S. zooepidemicus* bind low levels of IgG compared to other members of the *S. zooepidemicus* population (Lindmark *et al.*, 1999). The reasons for these phenotypic differences are unknown and may be numerous although variation in regulatory control might account for some. Eag/Zag are multiple ligand binding proteins that also bind the host protease inhibitor α_2-macroglobulin (α_2M) and albumin (Jonsson *et al.*, 1995). Binding of α_2M to cells of group C streptococci has been reported to inhibit phagocytosis *in vitro* (Valentin-Weigand *et al.*, 1990) possibly by protecting antiphagocytic surface structures from proteolytic degradation (Rasmussen *et al.*, 1999). Horses naturally infected with *S. equi* mount an antibody response to Eag confirming its production *in vivo* (Flock *et al.*, 2004). Immunisation with Eag as a subunit vaccine alone or in combination with 2 other recombinant proteins conferred partial protection against *S. equi* infection in mice and horses, respectively (Flock *et al.*, 2004; Flock *et al.*, 2006; Waller *et al.*, 2007).

Near identical proteins with homology to the Mac (IdeS) proteins of *S. pyogenes* were identified in *S. zooepidemicus* (IdeZ) and *S. equi* (IdeE) (Table 2.1) (Lannergard & Guss, 2006; Lei *et al.*, 2003). GAS Mac proteins bind to PMN Fc receptors and block antibody mediated opsonophagocytosis by these immune cells (Lei *et al.*, 2002a; Lei *et al.*, 2003). IdeE bound to neutrophils, particularly a subpopulation of these leukocytes that presumably expressed a distinct integrin and inhibited the phagocytosis of *S. equi* by PMNs (Timoney *et al.*, 2008). IdeE and IdeZ, like other streptococcal Mac proteins, also demonstrated IgG endopeptidase activity (Lannergard & Guss, 2006). Removal of surface bound IgG by this activity has been suggested as another potential mechanism for

avoidance of phagocytosis (von Pawel-Rammingen *et al.*, 2002). However, this function had little impact on Mac_{5005}-mediated resistance to phagocytosis (Lei *et al.*, 2002a) and IgG degradation may be a less critical feature of IdeE since *S. equi* is an obligate parasite of horses and IdeE failed to cleave equine IgG efficiently due to a lack of a distinct substrate site in the most prevalent equine subclass IgG4. These Mac proteins do not contain a wall anchor domain and are predicted to be secreted from the cell, although levels of IdeE/IdeZ detected in culture supernatants were slightly higher in *S. equi* compared to *S. zooepidemicus* and they also associated with the cell wall (Timoney *et al.*, 2008). When grown in the presence of serum, *S. equi,* but not *S. zooepidemicus,* degraded equine IgG (including IgG4) suggesting that an unidentified factor is produced by the former that contributes to the proteolysis of IgG. A continual removal of surface bound IgG through proteolysis is presumably an important housekeeping function to allow the pathogen to continue to bind and neutralise specific IgG capable of eliciting a complement cascade and phagocytosis (Soderberg & von Pawel-Rammingen, 2008).

The secreted *S. equi* unique Se18.9 significantly reduces the bactericidal activity of equine neutrophils toward this pathogen. The antiphagocytic properties of Se18.9 may depend, at least in part, on its ability to subvert the host's regulatory control of the complement cascade. Se18.9 binds to the complement regulator factor H, which serves as a cofactor for factor I-mediated cleavage of cell-bound C3b to inactive C3b and reduces the amplification of C3b via the alternative pathway. C3b deposition on *S. equi* was reduced following the addition of Se18.9 and increased when Se18.9 antiserum was added (Tiwari *et al.*, 2007).

Both *S. equi* and *S. zooepidemicus* direct a considerable proportion of carbon toward the production of a HA capsule which, due to the ubiquitous presence of HA in the ECM of vertebrates, is nonimmunogenic and serves to camouflage the bacteria from the host immune system. HA capsule confers resistance to phagocytosis in both subspecies (Anzai

et al., 1999b),(Wibawan *et al.*, 1999), (unpublished data) and, although not directly proven through gene inactivation, the presence of HA has been associated with virulence in *S. equi* (Anzai *et al.*, 1999b; Harrington *et al.*, 2002). Most *S. equi* isolates have a large colony phenotype on blood agar plates suggestive of a constitutive production of HA *in vitro*, whilst the recombinant *ΔhasA* mutant has a much smaller colony size (Carl Robinson, unpublished data). *S. zooepidemicus* strains however display considerable diversity in colony size and demonstrate phase variation in their degree of encapsulation (Soedarmanto *et al.*, 1996; Wibawan *et al.*, 1999).

Fne (see above) stimulates cell-mediated collagen gel contraction *in vitro* directed by the endothelial cell receptor αVβ3. By simultaneously binding collagen and fibronectin, Fne is thought to form a molecular bridge between collagen fibers and cell surface fibronectin that is in turn bound by αVβ3 integrin resulting in gel contraction (Liden *et al.*, 2008). As interstitial fluid pressure (IFP) can be lowered or normalised by a number of agents that also inhibit or stimulate collagen gel contraction *in vitro* respectively, this *in vitro* assay models IFP regulation during inflammation *in vivo*. Fne was shown to restore dermal IFP in mice, lowered as a result of anaphylaxis. Fne could therefore provide a novel virulence mechanism to protect *S. equi* from the innate immune response by inducing a compacted connective tissue, thereby counteracting a lowering of IFP and the resulting oedema formation that would usually enhance clearance of infectious organisms. This process may help to encapsulate *S. equi* infection and hinder the diffusion of macromolecules such as IgG, cytokines and inflammatory cells. The homologous N terminal portion of Fnz (FnzN) in *S. zooepidemicus* is also proven to mediate collagen gel contraction (Liden *et al.*, 2006), however, Fnz is proposed to be unlikely to have this activity because it is anchored to the bacterial cell wall (Liden *et al.*, 2008).

2.1.2.3 Tissue damage, dissemination and nutrient acquisition

Bacteria cannot survive and proliferate *in vivo* without access to nutrient sources, but the host has its own measures to deprive invading microorganisms of macro- and micronutrients. To counteract these defences bacteria use systems to sabotage host stores and actively take up nutrients. Many of the putative virulence factors already described (Eag/Zag), including those specific to *S. equi* (superantigens, SeM) could contribute to tissue breakdown in the host with a resultant release of valuable nutrients. Damage to the host might be thought of as a by-product of the organisms' nutrient-questing behaviour although improved invasion and dissemination are other possible consequences of cell degradation.

By binding α_2M via the multiple ligand binding proteins Zag and Eag, *S. zooepidemicus* and *S. equi* may protect their own surface structures from proteolytic degradation but a resultant depletion of surrounding protease inhibitors could also increase the amount of free proteases and enhance tissue destruction during the infection process (Rasmussen & Bjorck, 2002). Streptokinase, a secreted product of both subspecies activates the conversion of plasminogen to plasmin, a serine protease that can degrade fibrin clots and the ECM (McCoy *et al.*, 1991). The acquisition of superantigens by *S. equi* during its evolution from a *S. zooepidemicus* progenitor may have contributed to the more invasive nature of *S. equi* as a result of increased cell damage associated with an overzealous inflammatory response. Streptococcal exotoxins are implicated in severe group A streptococcal infections of man including scarlet fever and streptococcal toxic shock syndrome (epithelial damage leads to capillary leakage and hypotension) (Fraser & Proft, 2008). SeeH and SeeI enhanced the susceptibility of rabbits to lipopolysaccharide-induced endotoxic shock (Artiushin *et al.*, 2002). The *speL* gene has been recently acquired by M3/T3 *S. pyogenes* isolates in Japan (sometime from 1973 to 1992) that have caused more than 10% of streptococcal toxic shock-like syndrome cases reported from 1992 to 2002

(Ikebe *et al.*, 2002). In addition, the presence of the *speL* gene is highly associated with *S. pyogenes* M89, a serotype linked to outbreaks of rheumatic fever in New Zealand (Proft *et al.*, 2003).

S. *equi* and *S. zooepidemicus* also secrete a potent β-haemolytic toxin, characterised in *S. equi* to be streptolysin S-like (Flanagan *et al.*, 1998) (Table 2.1). Streptolysin S creates holes in the cell membrane of erythrocytes and many other mammalian cells resulting in their osmotic lysis and the release of nutrients, including haemoproteins which offer a rich source of iron to the invading pathogen. *S. equi* has a surface protein *Shp* and a lipoprotein *HtsA* that bind and transfer haem ready for its presumed uptake by the *htsABC* encoded ABC transporter (Nygaard *et al.*, 2006) (Table 2.1). Genes encoding this iron acquisition system are present in a conserved region of the *S. equi* genome relative to *S. pyogenes* and are shown from this study to exist in *Sz*H70 also.

A homologue of MtsABC, the manganese and iron transporter of *S. pyogenes* (Janulczyk *et al.*, 2003) has also been identified in *S. equi* (Harrington *et al.*, 2000) (Table 2.1). The gene encoding the lipoprotein component was identified in all of 10 disparate *S. equi* (n = 5) and *S. zooepidemicus* isolates (n = 5). The *S. equi* cell surface displays various protein components (HPr, LppC, Pst, HAP) that can either transfer, release or bind/transport phosphate groups at the extracytoplasmic surface, although their exact functions and contributions to virulence (if any) are unknown and may serve nutritional purposes by scavenging phosphate and organic phosphoesters and/or regulatory roles through environmental sensing (Dixon *et al.*, 2001; Hamilton *et al.*, 2000; Harrington *et al.*, 2002) (Table 2.1). Genes encoding these proteins are also present in *Sz*H70. HAP is hypothesised to play a role in HA capsule maintenance and was partially protective against *S. equi* and *S. zooepidemicus* infection in mice (Chanter *et al.*, 1999; Nickel *et al.*, 1998) but failed to prevent the development of strangles in ponies.

Prior to this study, very few differences have been highlighted in the capacity of *S. equi* and *S. zooepidemicus* to acquire carbon or nitrogen sources and essential trace elements. *S. zooepidemicus* is able to utilise lactose and sorbitol but *S. equi* lacks this ability, which allows these 2 subspecies to be differentiated biochemically. Carbohydrate catabolism has received more interest recently as an important participant in bacterial infection; specific mechanisms of carbohydrate usage appear to enhance the colonising potential of streptococci (Shelburne *et al.*, 2008a).

2.2 Aims

The overall aim of the work presented in this chapter was to investigate the genetic factors that separate *S. equi* and *S. zooepidemicus* in order to reveal the functional differences resulting from gene gain, gene loss and gene change that have contributed to the evolution of the more virulent, less versatile, host and niche-adapted pathogen *S. equi* from its presumed ancestor *S. zooepidemicus*. An applied objective was to identify potential vaccine and diagnostic targets, primarily to aid in the prevention and diagnosis of strangles. These aims were achieved using the following specific objectives:-

1. To identify genetic differences that delineate *Se*4047 and *Sz*H70 using the Artemis Comparison Tool (ACT).
2. To use bioinformatic and in some cases *in vitro* analysis to establish the function or putative function of proteins encoded by these novel gene loci.
3. To use PCR to establish the prevalence of some important genetic loci across a diverse selection of *S. equi* and *S. zooepidemicus* strains in order to clarify the true subspecies-specific genetic factors that have contributed to the specialization of *S. equi* (and/or subsets of the *S. zooepidemicus* population).

2.3 Methods

2.3.1 Bacterial strains

*Se*4047, ST-179 and SeM allele type 3, was isolated in 1990 from the submandibular abscess of a New Forest pony in Hampshire (Kelly *et al.*, 2006). *Sz*H70, ST-1, was isolated in 2000 from a Thoroughbred racehorse in Newmarket (Webb *et al.*, 2008). A set of 26 *S. equi* strains and 140 *S. zooepidemicus* strains representative of diverse SeM allele types and MLST STs, respectively were used for gene prevalence screening and functional analysis (Appendix Table A.1). The relatedness of MLST STs was determined by Katy Webb (AHT) and Keith Jolley (Oxford University) using ClonalFrame [58].

2.3.2 Identification of genetic differences between *Se*4047 and *Sz*H70

Genetic differences were identified following a pairwise comparison of the *Se*4047 and *Sz*H70 genomes by reciprocal TBLASTX analysis (cut off score of 100). The genomes were aligned and analysed using ACT (Carver *et al.*, 2005). Early investigation of the subspecies-specific DNA loci identified in the incomplete genomes of *Se*4047 and *Sz*H70 involved the use of ORPHEUS (Frishman *et al.*, 1998) and GLIMMER (Salzberg *et al.*, 1998) to aid the manual prediction of potential CDSs. The predicted protein sequences were searched against a non-redundant protein database using BLASTP and FASTA to assist the assignment of putative functions. Annotation was based, wherever possible, on characterised proteins and graphical display features of ACT were used to aid the identification of protein characteristics (assisted by the use of online tools such as SignalP (Emanuelsson *et al.*, 2007) and TMHMM http://www.cbs.dtu.dk/services/TMHMM-2.0/ accessed 18.11.08). Predicted proteins were also compared against the PFAM database of protein domain hidden Markov models (Bateman *et al.*, 2002). The completed genomes were later annotated by Matt Holden (Sanger Institute) as described previously (Parkhill *et al.*, 2000; Parkhill *et al.*, 2003). Orthologous proteins were identified as reciprocal best

matches using FASTA with subsequent manual curation. Orthology inferred from positional information was investigated using ACT. Pseudogenes had one or more mutations that would prevent correct translation; each of the inactivating mutations was subsequently checked against the original sequencing data. The sequence and annotation of the *Se*4047 and *Sz*H70 genomes are deposited in the EMBL database under accession numbers FM204883 and FM204884, respectively.

Sequences used by Matt Holden for comparative genomic analysis were: *S. zooepidemicus* MGCS10565 (CP001129) (Beres *et al.*, 2008), *S. uberis* 0140J (AM946015) (Ward *et al.*, 2009), *S. pyogenes* Manfredo (AM295007) (Holden *et al.*, 2007), *S. thermophilus* CNRZ1066 (CP000024) (Bolotin *et al.*, 2004), *S. suis* P1/7 (http://www.sanger.ac.uk/Projects/S_suis/), *S. pneumoniae* TIGR4 (AE005672) (Tettelin *et al.*, 2001), *S. sanguinis* SK36 (CP000387) (Xu *et al.*, 2007), *S. mutans* UA159 (AE014133) (Ajdic *et al.*, 2002), *S. agalactiae* NEM316 (AL732656) (Glaser *et al.*, 2002), *S. gordonii* str. Challis substr. CH1 (CP000725) (Vickerman *et al.*, 2007) and *Lactococcus lactis* subsp. *lactis* IL1403 (AE005176) (Bolotin *et al.*, 2001).

2.3.3 Gene prevalence studies

Strains were cultured from single colonies on COBA strep select plates (bioMérieux) and DNA was purified from colony material using GenElute spin columns according to manufacturer's instructions (Sigma). Gene prevalence was then determined by quantitative PCR (QPCR) using a SYBR green based method with a Techne Quantica instrument. For the QPCR, DNA diluted 1/100 was mixed with 0.3 µM forward and reverse primers (Table 2.2) and 1x ABsolute QPCR SYBR green mix (Abgene) in a total volume of 20 µl and subjected to thermocycling at 95°C for 15 min, followed by 40 cycles of 95°C for 15 s, 55°C for 30 s and 72°C for 30 s. Dissociation curves were analysed

following a final ramp step from 60°C to 90°C with reads at 0.5°C increments to rule out non-specific amplification. Positive, negative and no-template controls were included on each run. Data were analysed using Quansoft software (Techne). Crossing point values relative to those for the *gyrA* house-keeping gene were used to determine gene presence or absence. Supervised assistance was received for the set up of QPCR reactions by Karen Steward, Fern Ainslie and Melanie Osborne (AHT). Fischer's exact test was used to test the null hypothesis that there was no statistically significant difference in the prevalence of gene loci between *S. equi* subspecies or between *S. zooepidemicus* isolates associated with particular disease, tissue or gene types. Statistical significance was set at $P \leq 0.05$.

Table 2.2 Oligonucleotide primers used to screen diverse strains of *S. equi* and *S. zooepidemicus* for the presence of genes and a large inter-replichore inversion event, to conduct quantitative reverse transcription-PCR analysis and to PCR across the junction of a circular form of ICE*Se2*

Primer	Sequence (5'-3')	Gene / region
zm457	GGGTTAATGAGCCGATACTCTTTG	*lacE* (SZO15230)
zm458	CGTTCCTAATACCAAGCCAAGC	*lacE* (SZO15230)
18f	CGATGTAGCTAAGTTGGCTG	*rbsD* (SZO15140)
18r	AGGTGTATCCTCTCCTGTTC	*rbsD* (SZO15140)
28f	CGGGTAAGACAGTGATTGTC	*sorD* (SZO01750)
28r	ACCAAAATGCGCCCAACAGC	*sorD* (SZO01750)
66f	AACGCTTGTGGCTTTCTCC	*hysA* (SZO06680)
66r	ACTCTTTGCCCCTCTTGATC	*hysA* (SZO06680)
zm49	GACGAATTCAGAGTATGCGCGAATGCTAG	*srtC.2 – srtC.3* (SZO18280 - SZO18270)
zm50	GACGATATCACCAAGCATCATCAAAAGGTAG	*srtC.2 – srtC.3* (SZO18280 - SZO18270)
24f	GTATTTGGAACAGTAGCACCG	SZO08560
24r	CATCCTTAACATGCTTCGCC	SZO08560
zm492	CAGCAACCTTTCATTACCAAATTACC	*esaA*
zm493	AACAAAGACTTCAAATCAATCACAGC	*esaA*
zm432	GTGTGTTGATTATGGCCAGC	Conserved hypothetical protein CRISPR locus (SZO14370)
zm433	TATCAGAACGCCCTAGTGTC	Conserved hypothetical protein CRISPR locus (SZO14370)
1f	CAGATGATGTTCTAGAGATGGG	*slaA* (SEQ0849)
1r	CTCTAATAGCATCGGCTACG	*slaA* (SEQ0849)
zm381	CAAGTGCAACTATGGCAACAG	*slaB* (SEQ2155)
zm382	ATCCTGCCTTGAAATACTTTCG	*slaB* (SEQ2155)
zm455	TTAATACGGATGAAAATACTACAGTTTGG	*seeL* (SEQ1728)
zm456	CAACATTAAACATTCTTTCCTGTGAAG	*seeL* (SEQ1728)
4f	CTGTTAGGATGGTTTCTGCG	*seeM* (SEQ1727)
4r	TCAGCCGATAATGCAAGACC	*seeM* (SEQ1727)
15f	CAAGAGGCTTGTGAATGTCC	*seeH* (SEQ2036)
15r	CATGCTATTAAAGTCTCCATTGCC	*seeH* (SEQ2036)
15af	TTGGAGTATTCTCCTCCCTG	*seeI* (SEQ2037)
15ar	AGCATACTCTCTCTGTCACC	*seeI* (SEQ2037)
eqbE f	AAGATATAGCAGCATCGTATCG	*eqbE* (SEQ1242)
eqbE r	TCTAAATCTCTATTAAATAGCGGTATATTG	*eqbE* (SEQ1242)
zm340	ACTGATATTGAAGAAGCTTGAAG	*se18.9* (SEQ0235)
zm341	GCATAGGTGTCAGTCAAAGC	*se18.9* (SEQ0235)

Table 2.2 continued

Primer	Sequence (5'-3')	Gene / region
gyrA f	AAGGCGGGATTCCTAAAATC	*gyrA* (SEQ1170, SZO09430)
gyrA r	GATAAGTAAGCCCTCTAAAATGTG	*gyrA* (SEQ1170, SZO09430)
zm235	ATTGGGAACACCTTGCAAGG	across junction of circular ICE*Se2*
zm236	TTTTTTCTTCTTCCCACTGGC	across junction of circular ICE*Se2*
zm473F	TTAGTTTCTTTGCTTGTATTTGGAACAG	SZO08560
zm474R	TCTGCGTGTTTACTTCATCAGTATTATC	SZO08560
zm475R	CTTTTCTCTCATATATCCCTTTTCTTG	Recombinase (SZO08550)
zm476F	AGGAGCAAAGCTTCTTGATG	Recombinase (SZO08550)
zm490	CAGACAAGGTAGCCTAGTAC	across junction of large inter-replichore inversion in *Sz*H70
zm491	AACTCCTCATAGCACTCACG	across junction of large inter-replichore inversion in *Sz*H70

2.3.4 Sugar fermentation

The ability of isolates to ferment lactose, ribose and sorbitol was determined in Purple broth (Difco) using microtitre plates (ThermoScientific). Individual colonies were inoculated into Todd Hewitt broth (THB, Sigma), cultured overnight at 37 °C with 5% CO_2 and subsequently sub-cultured 1/5 into Purple broth (3.1%) supplemented with 1% sugar (Sigma) and 5% horse serum (Oxoid). Following a second overnight incubation at 37 °C with 5% CO_2, the ability to ferment the sugar tested was noted based on the presence or absence of a colour change from blue to yellow.

2.3.5 Mitogenicity assays

Mitogenicity assays were performed by Romain Paillot (AHT). Filter sterilized culture supernatants were prepared following the overnight culture of *S. equi* or *S. zooepidemicus* strains in THB (Sigma) at 37 °C with 5% CO_2. Equine PBMCs were purified from heparinised blood by centrifugation on a Ficoll density gradient. PBMCs were incubated with *S. equi* or *S. zooepidemicus* culture supernatants diluted 1/20. PBMC proliferation was detected by overnight incorporation of ^3H thymidine after 3 days of culture. Equine

PBMC proliferation is expressed as stimulation index (SI) calculated as follows (experimental response/control response). A SI $\geq$ 2 was considered as positive.

2.3.6 Reverse transcription and quantitative real-time PCR for recombinase activity

The potential for inversion of the promoter region downstream of the recombinase was assessed by comparison of SZO08560 mRNA transcript levels (produced when the promoter region is in the forward orientation) with reverse strand SZO08550 mRNA transcript levels (produced when the promoter region is inverted) in *Sz*H70. This analysis was performed by Karen Steward under my supervision.

Total RNA was extracted from bacteria using RNAprotect, RNeasy and DNase Kits according to the manufacturer's instructions (Qiagen). *Sz*H70 bacteria grown to log phase in THB (Sigma) with 10% horse serum (Oxoid) were pelleted by centrifugation for 10 min at 5000 g. The pellet was resuspended in 200 µl of TE buffer (10 mM Tris-Cl, 1 mM EDTA; pH 8.0) containing 3 mg lysozyme (Sigma) and 500 U mutanolysin (Sigma). After incubation at room temperature for 45 min with repeated vortexing, 700 µl of RLT buffer containing β-mercaptoethanol was added, and the tube was vortexed. The mixture was transferred to a 2 ml reaction tube containing 0.05 g of 100 µm-diameter acid-washed glass beads (Sigma) and vortexed for 5 min. The mixture was then centrifuged and total RNA was extracted from the supernatant using an RNeasy Midi Kit. 5 µg of recovered RNA was treated with DNase to remove any contaminating DNA followed by clean up on an RNeasy mini column. The quantity and purity of RNA was determined by optical density measurements at 260 and 280 nm using a NanoDrop^R ND1000 spectrophotometer (NanoDrop Technologies).

A quantitative two-step reverse transcription (RT) PCR procedure was used to analyse levels of SZO08560 and reverse strand SZO08550 transcription relative to the housekeeping gene *gyrA*. RT was performed using the Verso cDNA kit (Abgene). The RT reaction mixture (20 µl) contained 100 ng total RNA, 2 µM gene-specific primer (ZM474R or ZM476F) (Table 2.2), 500 µM dNTP mix, 1x cDNA synthesis buffer, 1 µl RT enhancer and 1 µl Verso enzyme mix. RT was performed at 50°C for 30 min and terminated by heating to 95°C for 2 min. QPCR was performed with a Techne Quantica instrument and data analysed using Quansoft software (Techne). For the QPCR, 6 µl RT reaction mixture diluted 1/1000 was mixed with 0.3 µM forward and reverse primers (Table 2.2), and 1x ABsolute QPCR SYBR green mix (Abgene) in a total volume of 20 µl and subjected to thermocycling at 95°C for 15 min, followed by 40 cycles of 95°C for 15 s, 55°C for 30 s and 72°C for 30 s. Dissociation curves were analysed, following a final ramp step from 60°C to 90°C with reads at 0.5°C increments, to rule out non-specific amplification. No-template negative controls were included and reverse transcriptase negative controls to confirm the absence of contaminating DNA from RNA samples. Standard curves (Crossing point (Cp) vs. log gene copy number) were generated from genomic DNA for each target gene and used to calculate transcript copy number in cDNA samples. SZO08560 and reverse strand SZO08550 transcript copy numbers were normalised to *gyrA* reference gene copy number to correct for differences in the amount of starting material. Data were expressed as fold difference in normalised SZO08560 transcript level relative to reverse strand SZO08550 transcript level.

2.4 Results

2.4.1 Genome alignment of *Se*4047 and *Sz*H70 using the Artemis Comparison Tool: identification and prevalence of novel gene loci

2.4.1.1 Overview

MLST has provided evidence of the close genetic relationship of *S. equi* and *S. zooepidemicus* (Webb *et al.*, 2008). The genomes of *Se*4047 (ST-179) and *Sz*H70 (ST-1) support the overall relatedness, but also reveal evidence of genome plasticity that has generated notable diversity. The two genomes are similar in size: the *Se*4047 genome consists of a circular chromosome of 2,253,793 bp (Figure 2.1A) encoding 2,137 predicted CDSs, and the *Sz*H70 genome contains a chromosome of 2,149,866 bp (Figure 2.1B), encoding 1,960 predicted CDSs (Holden, 2009). Much of the *Se*4047 genome is orthologous to the *Sz*H70 genome: 1671 *Se*4047 CDSs have *Sz*H70 orthologs. Of the remaining 466 non-orthologous *Se*4047 CDSs, 422 are found on mobile genetic elements (MGEs) (Holden, 2009). For details of the regions of variation in the *Se*4047 and *Sz*H70 genomes see Appendix Table A.2.

Recently, the genome sequence of *S. zooepidemicus* strain MGCS10565 (*Sz*MGCS10565) was published (Beres *et al.*, 2008). This strain was isolated from a human case of nephritis that was part of a severe epidemic in Brazil (Balter *et al.*, 2000). MLST (http://pubmlst.org/szooepidemicus/) analysis indicates that *Sz*H70 and *Sz*MGCS10565 (ST-72) are genetically distinct from each other and *Se*4047. Comparative analysis reveals that the number of orthologs in the *Se*4047 genome is slightly higher for *Sz*H70 (78.2%) than for *Sz*MGCS10565 (77.4%); 76.3% of the *Se*4047 CDSs have matches in both *S. zooepidemicus* strains (Holden, 2009). The following analysis has primarily focused on the comparison of the equine isolates, *Se*4047 with *Sz*H70 but some details of *Sz*MGCS10565 are included for comparison.

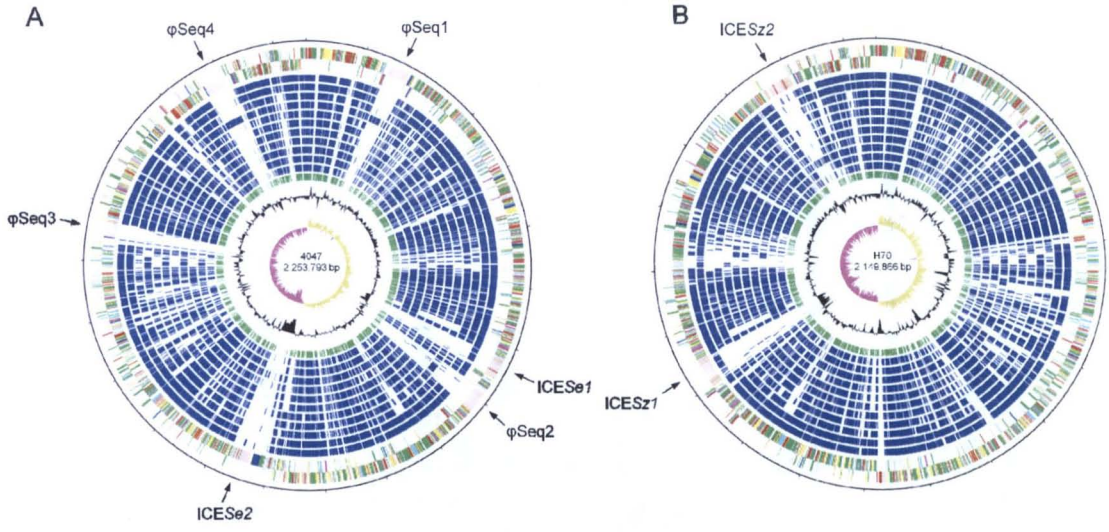


Figure 2.1 Schematic circular diagrams of the *Se*4047 (A) and *Sz*H70 genomes (B)

Figure reproduced with kind permission from Matt Holden, Sanger Institute (Holden, 2009). Key for the circular diagrams (outside to inside): scale (in Mb); annotated CDSs coloured according to predicted function represented on a pair of concentric circles, representing both coding strands; orthologue matches shared with the streptococcal species, *Se*4047 or *Sz*H70, *Sz*MGCS10565, *S. uberis* 0140J, *S. pyogenes* Manfredo, *S. mutans* UA159, *S. gordonii* Challis CH1, *S. sanguinis* SK36, *S. pneumoniae* TIGR4, *S. agalactiae* NEM316, *S. suis* P1/7, *S. thermophilus* CNRZ1066, blue; orthologue matches shared with *Lactococcus lactis* subspecies *lactis*, green; G + C% content plot; G + C deviation plot (> 0% olive, < 0% purple). Colour coding for CDS functions: dark blue; pathogenicity/adaptation, black; energy metabolism, red; information transfer, dark green; surface associated, cyan; degradation of large molecules, magenta; degradation of small molecules, yellow; central/intermediary metabolism, pale green; unknown, pale blue; regulators, orange; conserved hypothetical, brown; pseudogenes, pink; phage and IS elements, grey; miscellaneous. The positions of the 4 prophage and 2 ICE*Se* present in the *Se*4047 genome, and 2 ICE*Sz* in the *Sz*H70 genome are indicated.

Comparison of the predicted functions of the proteins encoded in the *Se*4047 and *Sz*H70 genomes revealed that *Se*4047 has the same number, or fewer CDSs, in each of the functional classes with the exception of protective responses and adaptation and laterally acquired elements (Figure 2.2A) (Holden, 2009). The number of pseudogenes in *Se*4047 is also elevated in comparison to *Sz*H70 (Figures 2.2A and 2.2B).

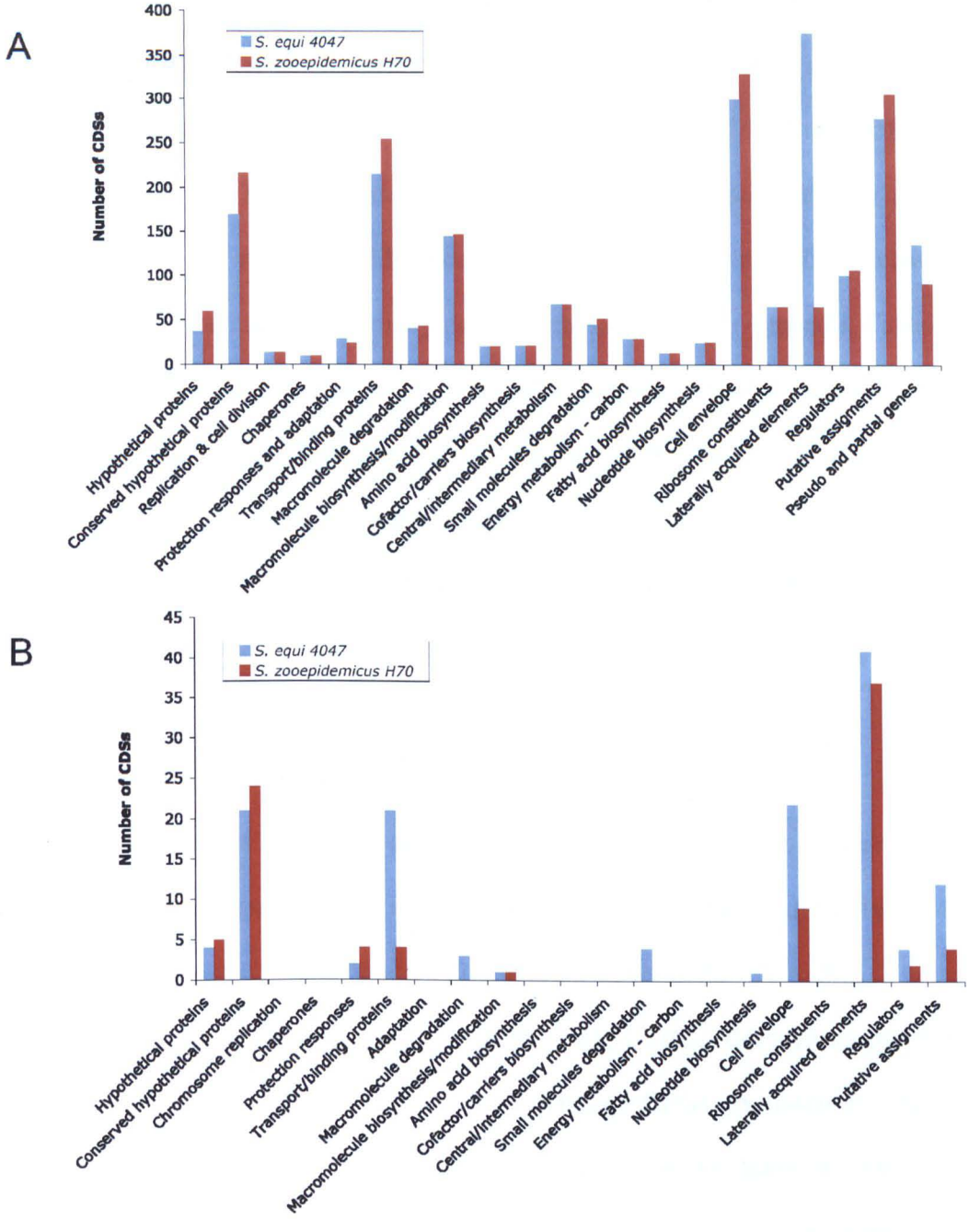


Figure 2.2 Distribution of CDSs belonging to different functional classes in the *Se*4047 and *Sz*H70 genomes

Figure reproduced with kind permission from Matt Holden, Sanger Institute (Holden, 2009). (A) Functional CDSs of *Se*4047 and *Sz*H70. (B) Partially deleted or pseudogenes in the *Se*4047 and *Sz*H70 genomes.

Closer examination of the genomes revealed (excluding IS elements) 31 and 37 novel genetic loci encoding 336 and 247 predicted proteins in *Se*4047 and *Sz*H70, respectively

when compared to each other (Appendix Tables A.2A and A.2B, Figure 2.3). At least 80% of the *Se*4047-specific CDSs have been gained through the integration of 4 prophages (φSeq1 – 4), an ICE (ICE*Se2*), and 2 other genetic loci (Figures 2.1A and 2.4A), whilst deletion events in *Se*4047 relative to *Sz*H70 account for at least 35% of the CDSs specific to the latter (Figure 2.4B). Loss of function through gene inactivation is also more prevalent in *Se*4047, which has 136 pseudogenes (including 58 partially deleted genes) compared to 91 (including 62 partially deleted genes) in *Sz*H70 (Appendix Tables A.2D and A.2E). In the shared core genome, 72% of pseudogenes identified in *Se*4047 are intact in *Sz*H70, while 57% of *Sz*H70 pseudogenes are also dysfunctional in *Se*4047 (Appendix Tables A.2D and A.2E). In particular, *Se*4047 is enriched for mutations associated with catabolic metabolism, transport, and the cell envelope (Figure 2.2B) (Holden, 2009). Overall, these data suggest that the evolution of *S. equi* has been shaped by gene gain and loss.

*Sz*H70 has acquired its own set of novel loci through gene gain when compared with the *Se*4047 genome, a feature that is likely to reflect the genetic diversity in the *S. zooepidemicus* population and the more distant relationship of this ST to *S. equi* compared with other *S. zooepidemicus* strain types (Webb *et al.*, 2008). Although no prophage are present in *Sz*H70 (or *Sz*MGCS10565), 18% of its unique CDSs (excluding ISs) have been gained through the insertion of a novel ICE (ICE*Sz2*) and one genomic island (Figures 2.1B and 2.4B). Many of the novel DNA loci have atypical GC content (Appendix Table A.2) compared to the genome as a whole (*Se*4047: 41.28%; *Sz*H70: 41.53%), but only those that encode a putative integrase and/or are flanked by duplicated repeat elements have been referred to as genomic islands. In a number of cases it was not possible to ascertain whether insertion or deletion events were responsible for the gain or loss of CDSs in one genome relative to the other (Figure 2.4). However, Figure 2.4A demonstrates the low occurrence of obvious deletion events in *Sz*H70 when compared to *Se*4047.

One ICE is shared between the 2 genomes (ICE*Se1* and ICE*Sz1*) but these elements in *Se*4047 and *Sz*H70 have a mosaic structure relative to one another and display variation in their N-terminal and C-terminal regions possibly due to recombination events. A number of other regions shared between *Se*4047 and *Sz*H70 show significant divergence in their number and/or sequence of genes (Appendix Tables A.2A-C).

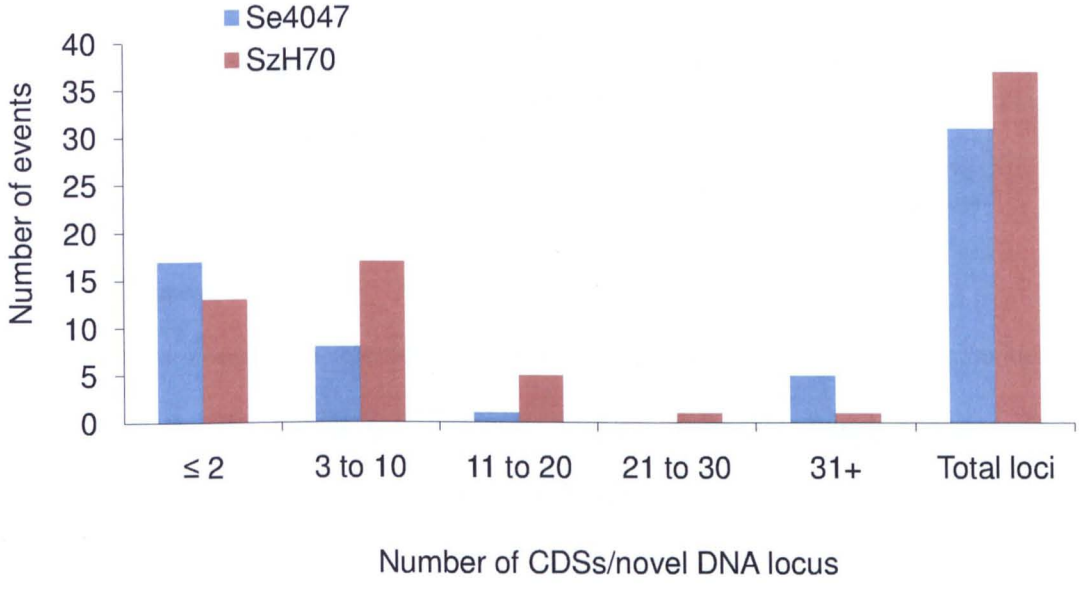


Figure 2.3 Distribution of subspecies-specific DNA loci

Bar chart shows the number of CDSs present on DNA loci that are novel to *Se*4047 or *Sz*H70 and the total number of novel DNA loci in each subspecies.

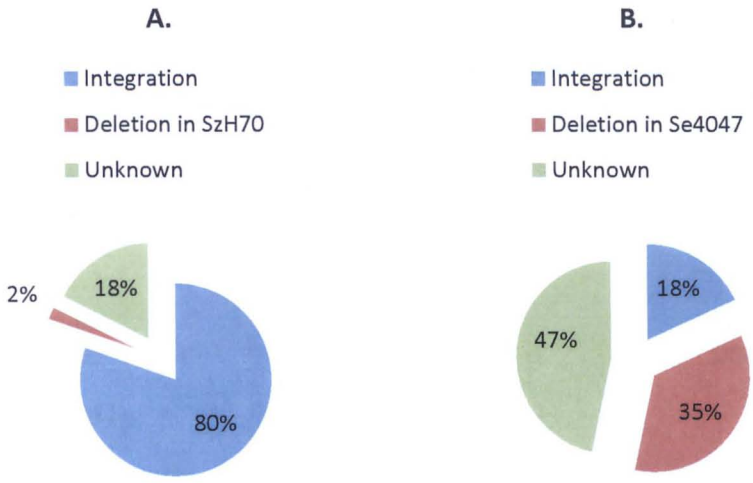


Figure 2.4 Pie charts showing the proportion of species-specific CDSs (not including ISs) in (A) *Se*4047 and (B) *Sz*H70 that are present through insertion, deletion or unknown events in each genome relative to the other

2.4.1.2 IS elements

Both genomes contain IS elements of 5 or more families, although a considerable expansion of the IS3 family is a particular feature of *Se*4047, which contains 73 ISs compared to 30 in *Sz*H70 (Table 2.3) (Holden, 2009).

Table 2.3 Composition of insertion sequences in the *Sz*H70 and *Se*4047 genomes

Insertion element	*Sz*H70	*Se*4047	Family
ISSzo1 (ISSeq1)	4	4	ISL3
ISSzo2 (ISSeq2)	5	3	ISNCY
ISSzo3 (ISSeq3)	4	40	IS3
ISSzo4 (ISSeq4)	5	11	IS1634
ISSzo5 (ISSeq5)	2	8	IS3
ISSzo6	1	0	IS30
ISSzo7	3	0	IS3
ISSzo8 (ISSeq8)	2	2	IS200/IS605
ISSzo9 (ISSeq9)	1	1	IS200/IS605
ISSzo10	2	0	ISAs1
ISSzo11 (ISSeq11)	1	1	ISL3
ISSeq14	0	1	ISL3
ISSeq15*	0	1	IS3
Total	30	73	

Table reproduced with kind permission from Matt Holden, Sanger Institute (Holden, 2009). IS elements were grouped into separate isoforms (IS elements with DNA sequence more than 95% identity), numbered accordingly, and given a specific 3 letter identifier to designate the species of origin. Truncated IS elements lacking either the 3-prime or 5-prime ends were not included in the table. *chimeric IS element that appears to have been generated from recombination between ISSeq3 and ISSeq5 elements.

2.4.1.3 Genome rearrangements

The *Se*4047 and *Sz*H70 chromosomes are generally collinear except for 2 inversions around the origin of replication (Figure 2.5). The smaller central inversion (spanning 119 kb) is due to recombination events in *Se*4047 between identical conversely orientated ISSeq3 elements on opposite replichores. This inversion event relocates part of a bacteriocin gene cluster (SEQ 1226 – SEQ1223) and an ABC transporter CDS (SEQ1276), both of which were already disrupted by insertion of the ISSeq3 elements and other DNA sequences (including ICE*Se2*). The larger rearrangement is due to an inter-replichore inversion in *Sz*H70 of unknown origin. 2 parts of a putative nudix gene that had already been disrupted by the insertion of a novel ICE (ICE*Sz2*), have been separated by this inversion event. PCR across one of the flanking sequences suggests a similar reorganization in 21% of *S. zooepidemicus* isolates (n = 140, Figure 2.6), although this inversion has not occurred in the genome of *Sz*MGCS10565. The inversion was not detected in any isolate of *S. equi* tested by PCR (n = 26), which highlights a statistically

significant difference in the genome organisation of these 2 subspecies (P = 0.0088, Fischer's exact test). Both the *Se*4047 and *Sz*H70 genomes contain two copies of *hasC* which encode UDP-glucose pyrophosphorylases (Blank *et al.*, 2008). In *Sz*H70 one copy of *hasC* (SZO17510) has been translocated to the opposite replichore by the previously mentioned large reciprocal inversion. There is also a small intra-replichore inversion (14 kb) in *Se*4047 between the 2 copies of *hasC* (SEQ0271 and SEQ0289), which rearranges the genes associated with capsule production (Blank *et al.*, 2008).

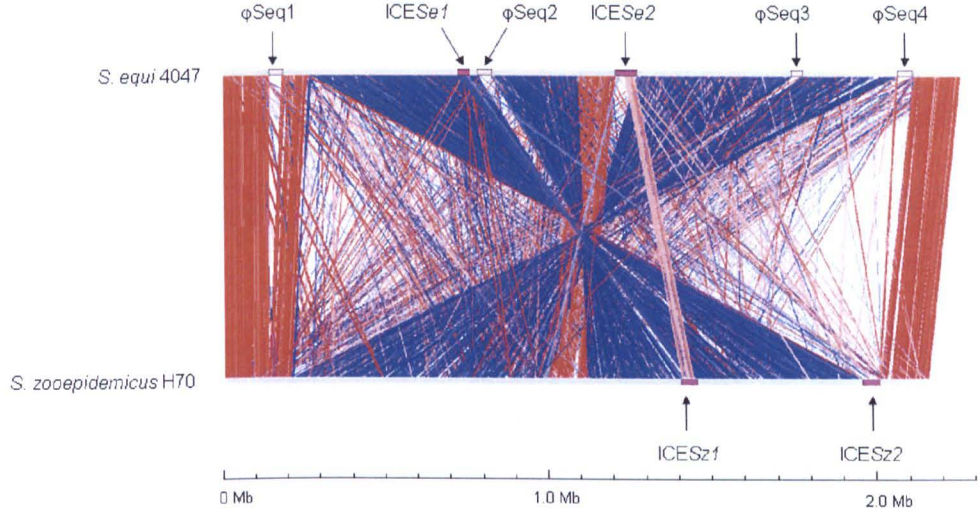


Figure 2.5 Pairwise comparison of the chromosomes of *Se*4047 and *Sz*H70 using ACT

Figure reproduced with kind permission from Matt Holden, Sanger Institute (Holden, 2009). The sequences have been aligned from the predicted replication origins (oriC; right). The coloured bars separating each genome (red and blue) represent similarity matches identified by reciprocal TBLASTX analysis, with a score cut off of 100. Red lines link matches in the same orientation; blue lines link matches in the reverse orientation. The prophage (pink) and ICE (purple) are highlighted as coloured boxes.

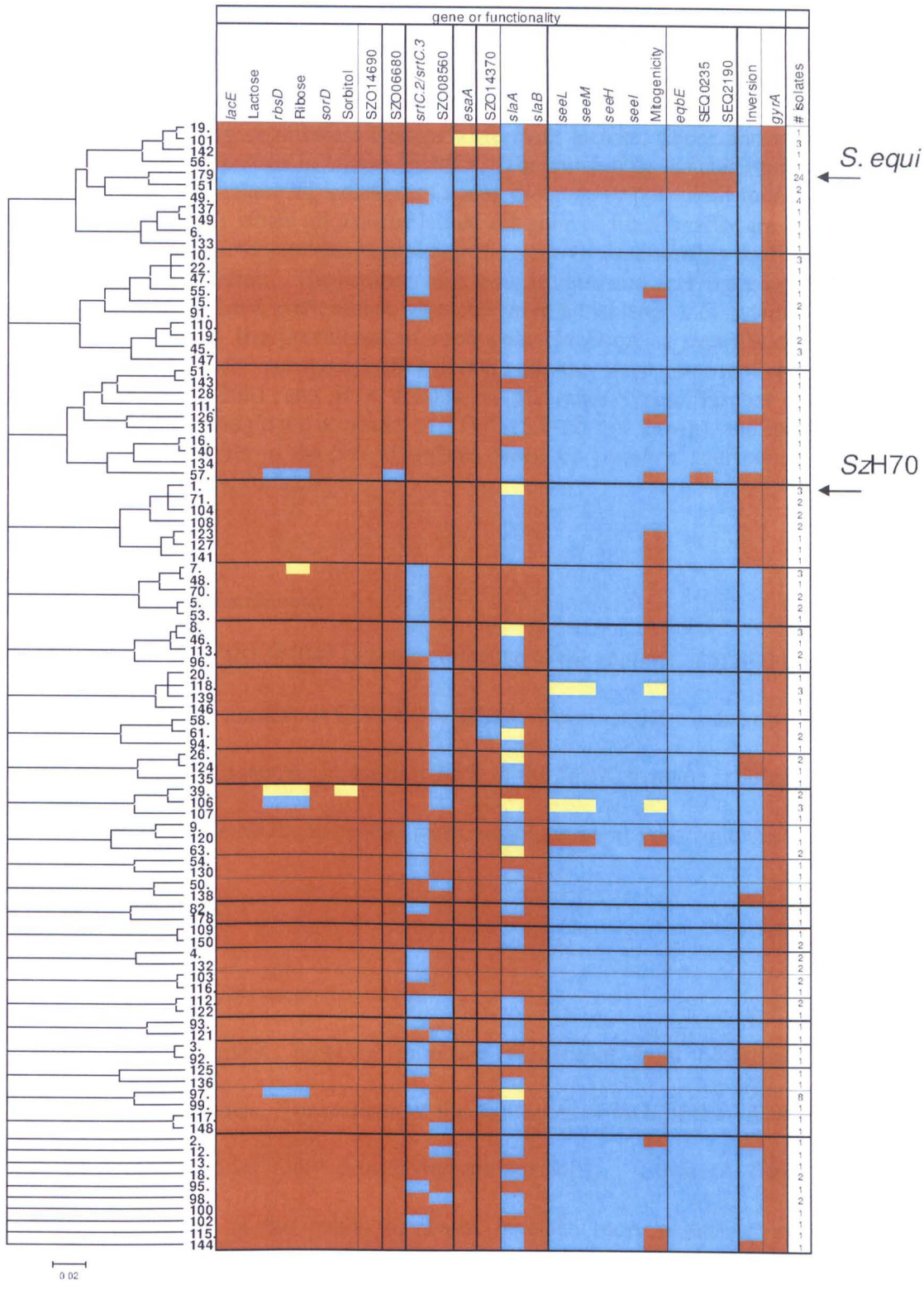


Figure 2.6 ClonalFrame analysis of MLST alleles of 26 *S. equi* and 140 *S. zooepidemicus* isolates and its relationship with the prevalence of selected differences between the *Se*4047 and *Sz*H70 genomes

The ClonalFrame tree (on the left) is reproduced with kind permission from Keith Jolley, Oxford University. Genes examined were *lacE*, *rbsD*, *sorD*, SZO14690 (encoding a putative branched chain amino acid transporter), SZO06680 (encoding a putative HA lyase and specific to the 4 bp missing from SEQ1479), *srtC.2/srtC.3*, SZO08560 (encoding

Listeria-Bacteroides repeat Pfam domains (PF09479)), *esaA*, SZO14370 (within the CRISPR locus), *slaA*, *slaB*, *seeL*, *seeM*, *seeH*, *seeI*, *eqbE* (within the equibactin locus), SEQ0235 (encoding Se18.9), SEQ2190 (encoding a cell-wall anchored protein), inversion (large inter-replichore inversion SZO01800 - SZO18031) and *gyrA*. Functional assays determined the ability of different isolates to ferment lactose, ribose and sorbitol and to induce mitogenic responses in equine PBMCs. The number of isolates representing each ST is indicated. STs where all isolates contained the gene or possessed functional activity are shown in red, STs where all isolates lacked the gene or functionality are shown in blue and STs containing some isolates containing the gene or functionality and some that did not are coloured in yellow. The position of *S. equi* isolates and *Sz*H70 are indicated. With the exception of *slaB* and *gyrA*, which were present in all isolates of *S. equi* and all isolates of *S. zooepidemicus*, the prevalence of each gene, function or event was significantly different between the 2 subspecies ($P = 0.0088$ for the large inter-replichore inversion SZO01800 - SZO18031 and $P < 0.0001$ in all other cases, Fischer's exact test). *Sz*MGCS10565 is a single locus variant of ST-10 (ST-72; not shown), and had an identical gene prevalence profile to the ST-10 isolates based on *in silico* analysis of its genome sequence (Beres *et al.*, 2008).

2.4.1.4 Mobile genetics elements

The increased size of the *Se*4047 genome compared to the genome of *Sz*H70 is due to the acquisition of a large number of CDSs from MGEs. Together these make up a total of 16.4% of the *Se*4047 genome. In contrast 7.5% of the *Sz*H70 genome is composed of MGEs. Several of the MGEs in *Se*4047 carry notable virulence determinants absent in *Sz*H70.

2.4.1.4.1 Prophages and associated exotoxins

*Se*4047 is polylysogenic, however prophage are absent from both the *Sz*H70 and *Sz*MGCS10565 genomes. Comparison of the sequences of each of the 4 *Se*4047 prophage with each other showed only limited mosaic similarity. However, comparison with prophage sequences in the public databases revealed more extensive similarity with prophage from *S. pyogenes*. Clustering analysis has demonstrated that the individual *S. equi* prophage are more related to phage in the other sequenced *S. pyogenes* genomes than they are to each other (Holden, 2009), suggesting commonality in the phage pool of these pathogens.

The first of the 4 *Se*4047 prophage, φSeq1, is 39 kb in size, contains the CDSs SEQ0133-SEQ0197 and is integrated immediately after the tRNA-Cys gene. The CDSs of φSeq1 do not have homology to known virulence factors. However, examination of the *Se*4047 genome sequencing reads revealed evidence that induction of φSeq1 prophage may occur (Nicholas Thompson, Sanger Institute). Re-circularised φSeq1 has since been amplified by PCR following phage induction by mitomycin C treatment and preparation of phage particles present in cultures of *Se*4047 (Holden, 2009).

The 41 kb φSeq2 (SEQ0787-SEQ0851) is integrated into the putative C-terminal sequence of an ATP-dependent DNA helicase (SEQ0786) and contains a CDS (SEQ0849) that shares 98% predicted amino acid sequence identity with the phospholipase A_2, SlaA, of *S. pyogenes* M3 MGAS315 (Beres *et al.*, 2002). *slaA* was present in all 26 strains of *S. equi* and 44 out of 140 *S. zooepidemicus* isolates tested by QPCR (Figure 2.6). This difference in prevalence of *slaA* between the 2 subspecies was statistically significant ($P < 0.0001$) using Fischer's exact test. Another homologue of *slaA*, denoted *slaB*, was identified at a different genome location in both *Se*4047 (SEQ2155) and *Sz*H70 (SZO18670), adjacent to a single prophage gene (SEQ2154/SZO18660). SlaB shares 70% amino acid sequence identity with SlaA of *S. pyogenes* and was present in all strains of *S. equi* and *S. zooepidemicus* examined (Figure 2.6).

The 30 kb φSeq3 is integrated into SEQ1725 (a putative late competence protein CDS) and contains CDSs SEQ1727-SEQ1765 including two cargo CDSs encoding the superantigens SeeL and SeeM, which share 97% and 96% amino acid sequence identity with SpeL and SpeM of *S. pyogenes* MGAS8232, respectively (Alber *et al.*, 2005; Proft *et al.*, 2003). *seeL* and *seeM* were present in all strains of *S. equi* and 4 isolates (ST-106, ST-118, ST-120) of *S. zooepidemicus* (of 140 isolates examined) that were recovered from the same outbreak of respiratory disease in 1996 (Figure 2.6).

Finally, the 40 kb φSeq4 is inserted into SEQ2035, resulting in the truncation of this putative transcriptional repressor. φSeq4 contains cargo CDSs encoding the previously described superantigens SeeH (SEQ2036) and SeeI (SEQ2037), which share 98% and 99% amino acid sequence identity with SpeH and SpeI, respectively (Artiushin *et al.*, 2002). φSeq4 is very closely related to φMan3 of *S. pyogenes* Manfredo (Figure 2.7, Matt Holden). Although *seeH* and *seeI* were present in all strains of *S. equi* tested, no strains of *S. zooepidemicus* have yet been shown to contain these genes.

To provide functional evidence for the production of superantigens by different strains of *S. equi* and *S. zooepidemicus*, Romain Paillot (AHT) assayed the culture supernatants of our strain collection. This analysis confirmed that all strains of *S. equi* and the strains of *S. zooepidemicus* containing *seeL* and *seeM* possessed significant mitogenic activity. However, the supernatants of 25 additional strains of *S. zooepidemicus* also had mitogenic activity. Several of these strains were related genetically by MLST, and clustered into three related groups (ST-123, ST-127 and ST-141; ST-7, ST-48, ST-70, ST-5 and ST-53; ST-8, ST-46 and ST-113) (Figure 2.6). Overall, mitogenic activity was exhibited by a significantly higher proportion of *S. equi* isolates (100%, n = 26) compared to *S. zooepidemicus* isolates (21%, n = 140, $P < 0.0001$, Fischer's exact test). In addition, mitogenic activity was significantly more prevalent in respiratory isolates of *S. zooepidemicus* (30%, n = 73) compared to other *S. zooepidemicus* isolates (9%, n = 65, P = 0.0027, Fischer's exact test).

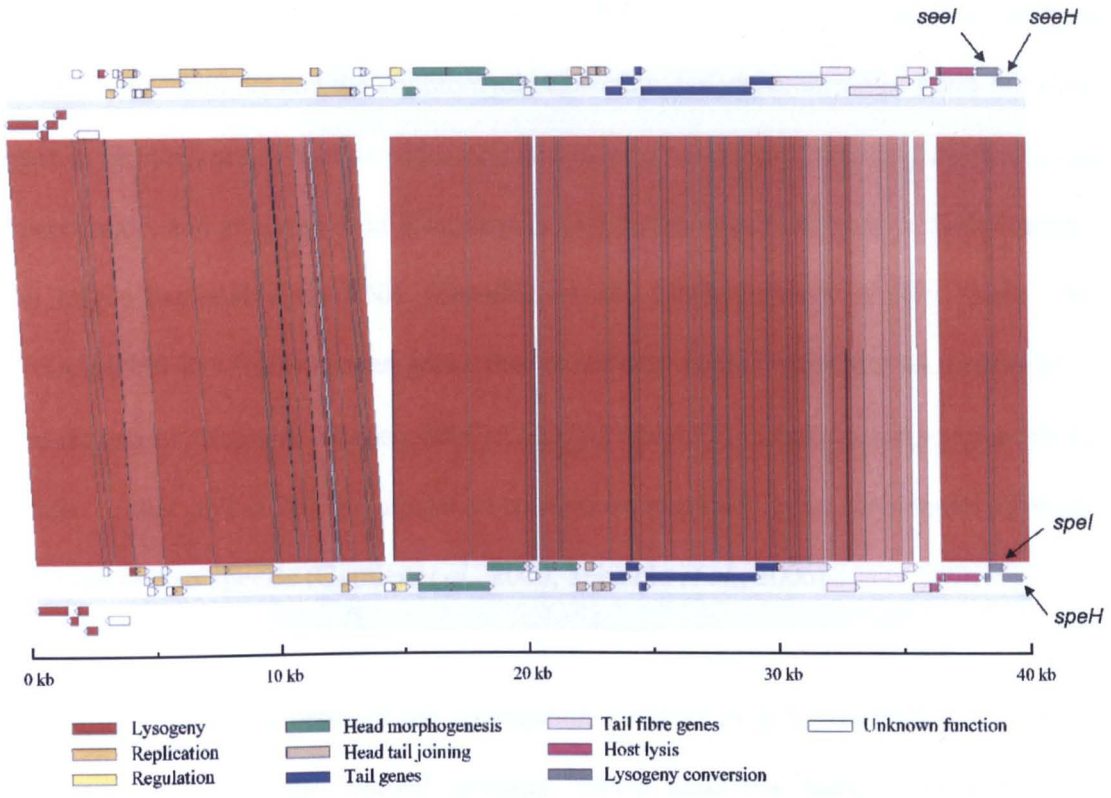


Figure 2.7 Pairwise comparison of *Se*4047 φSeq.4 and φMan.3 from *S. pyogenes* Manfredo displayed using ACT

Figure reproduced with kind permission from Matt Holden, Sanger Institute. *Se*4047 φSeq.4 is shown at the top of the alignment and φMan.3 from *S. pyogenes* Manfredo is shown at the bottom of the alignment. The red bars separating each sequence represent similarity matches identified by TBLASTX analysis. The locations of *seeI*, *seeH*, *speI* and *speH* are indicated.

2.4.1.4.2 Integrative and conjugative elements (ICEs)

Both the *Sz*H70 and *Se*4047 genomes contain distinct ICE regions (Appendix Table A.2). ICE*Se1* (SEQ0730a to SEQ0763) and ICE*Sz1* (SZO12981 to SZO12560) are homologous ICE elements of the *Se*4047 and *Sz*H70 genomes, respectively, and are integrated adjacent to a putative RNA methyltransferase gene (SZO12990/SEQ0729). These ICEs display heterogeneity in their N- and C-terminal regions, with a novel putative modification methylase (SEQ0757) and type-II restriction enzyme (SEQ0758) encoded by ICE*Se1*. BLAST analysis showed that ICE*Se1* and ICE*Sz1* share most homology with the conjugation modules of CDTn*2* and CDTn*5* in *C. difficile* (Sebaihia *et al.*, 2006), and also some homology with the conjugation CDSs of ICE MGAS10750-RD.2 of *S. pyogenes*

strain MGAS10750 (Beres & Musser, 2007). These ICEs carry transporters of unknown function in *C. difficile* and the erythromycin ribosome dimethyltransferase *ermA* (*erm(TR)*) gene in *S. pyogenes*, which provides co-resistance to macrolide (including erythromycin), lincosamide, and streptogramin B antibiotics (MLS_B resistance) through a modification of the target bacterial 23S rRNA (Sebaihia *et al.*, 2006; Seppala *et al.*, 1998). ICE MGAS10750-RD.2 also carries genes that potentially confer resistance to spectinomycin and tetronasin (Beres & Musser, 2007). CDTn*2* and CDTn*5* have a conjugation module related to that of Tn*1549*, a conjugative transposon responsible for vancomycin resistance in *Enterococcus* species (Garnier *et al.*, 2000; Sebaihia *et al.*, 2006).

Similarities also exist between the conjugation modules of ICE*Se1*/ICE*Sz1* and ICE*Se2*, the second ICE in the *Se*4047 genome, which also has features related to ICE MGAS10750-RD.2 (Figure 3.3, Chapter 3), CDTn*2*/CDTn*5* and Tn*1549* in *Enterococcus* spp (Figure 3.2, Chapter 3). However, ICE*Se2* encodes a unique siderophore-like non-ribosomal peptide synthesis system responsible for the production of a secreted product involved in iron acquisition (Chapter 3). A biosynthetic gene from this region of ICE*Se2* was present in all of the *S. equi* isolates, but in none of the diverse collection of *S. zooepidemicus* isolates examined by QPCR (Figure 2.6). A putative bacteriocin locus is also present outside of the predicted conjugation module in ICE*Se2*. An upstream putative two-component sensor histidine kinase is truncated at its N terminus, which may affect the regulatory control of this bacteriocin gene cluster. ICE*Se2* is integrated adjacent to an IS3 element, which is associated with a 119 kb inter-replichore inversion in the centre of the *Se*4047 chromosome. 108mer flanking inverted repeats of ICE*Se2* extend into this IS and into a truncated hypothetical protein at the other end of the element. This hypothetical protein (SEQ1273) and 2 other CDSs (SEQ1274, SEQ1274) are also novel to the *Se*4047 genome. ICE*Se2* was present in the same genome location in all *S. equi* strains tested (Chapter 3).

A unique ICE in *Sz*H70 (ICE*Sz*2: SZO18030 to SZO17561), which has integrated into a putative nudix hydrolase (SZO18031-SZO01811), is related to regions of probable ICEs in the genomes of *S. agalactiae* (NC_004116.1, NC_004368.1) and *S. suis* (NC_009442.1, NC_009443.1). ICE*Sz*2 encodes a number of proteins that may offer a selective advantage to this strain, including a multi-antimicrobial extrusion (MATE) family transporter (SZO17690), a putative glucan-binding protein (SZO17800) and a C-5 cytosine-specific DNA methylase (SZO17990). ICEs are absent from the genome of *Sz*MGCS10565 (Beres *et al.*, 2008).

2.4.1.5 A clustered regularly interspaced short palindromic repeat (CRISPR) array

A 23 kb region of the *Sz*H70 genome has been deleted in *Se*4047 through recombination between ISSeq11 elements (Table 2.3, Appendix Table A.2). 24 CDSs are present in this part of the *Sz*H70 genome including 7 CRISPR-associated genes (SZO14330-SZO14470) and a CRISPR array (Figure 2.8).

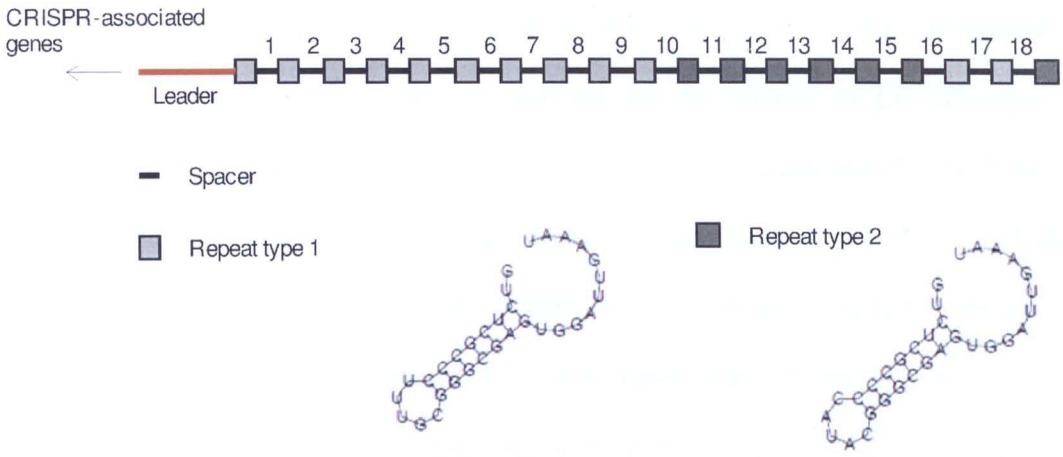


Figure 2.8 The clustered, regularly interspaced short palindromic repeat (CRISPR) locus of *Sz*H70

An AT-rich leader sequence separates the short palindromic repeat sequences interspersed with unique spacer sequences from the putative CRISPR-associated genes. The predicted secondary structure of the type 1 and type 2 repeats is shown using RNAfold (Gruber *et al.*, 2008).

CRISPR systems are a recently discovered defence mechanism, present in about 40% of sequenced bacterial genomes, which confer acquired phage resistance to bacteria (Kunin *et al.*, 2007; Sorek *et al.*, 2008). The *Sz*H70 CRISPR array is composed of 19 direct palindromic repeats, each 32 bp in size, separated by similarly-sized non-repetitive spacers (18 in total). Following transcription into RNA, these repeats are hypothesised to form RNA stem-loop structures (Sorek *et al.*, 2008). RNAfold software was used to predict the secondary structure of the 2 variants of these direct repeats that contain a 7 bp palindrome and differ by 3 nucleotides outside of the complementary palindromic region (Figure 2.8). The structural features of the RNA molecule and a conserved 3' terminus have been suggested to act as binding sites for one or more of the CRISPR-associated proteins. In this instance the 3' terminus motif is GAAAU, although the consensus varies slightly depending on which phylogenetic cluster the CRISPR repeat belongs to (Kunin *et al.*, 2007). Small RNA species derived following transcription of the spacer region (and repeat

region) are thought to base pair with foreign nucleic acids, as directed by the sequence of the spacer region, leading to their degradation (by the activity of CRISPR-associated protein(s)) (Sorek *et al.*, 2008). All 18 *Sz*H70 spacers are unique and BLASTN analysis revealed full length (or near full length) matches to prophage sequences in 7 of the 18 spacers (Table 2.4). 2 share > 94% identity with prophage sequences present in the published genomes of *S. pyogenes*, 4 spacers have identical matches with prophage sequences found in the *Se*4047 genome (#6 with SEQ0163 of φSeq1, #7 with SEQ1743 of φSeq3, #8 with SEQ1745 of φSeq3 and #15 with SEQ1727 (*seeM* φSeq3) and one spacer (#18) has a near identical match with the *Se*4047 prophage CDSs SEQ0190 (φSeq1) and SEQ1729 (φSeq3), differing only at the first nucleotide (C to T).

The CRISPR-associated genes of the *Sz*H70 CRISPR locus are present in the CRISPR 1 locus of *Sz*MGCS10565 and the leader sequences that separate the CRISPR-associated genes from the CRISPR array in both genomes differ by only 8 bp. A single type 2 *Sz*H70 repeat and 9 type 1 *Sz*H70 repeats are also present in CRISPR 1 of *Sz*MGCS10565 but 8 of the 9 interspersed spacers have unique sequences in comparison to the *Sz*H70 CRISPR. However, spacer 9 of *Sz*MGCS10565 CRISPR 1 has an exact match with spacer 18 of *Sz*H70 (Beres *et al.*, 2008). Another 2 CRISPR arrays with distinct repeats and spacer sequences were identified in *Sz*MGCS10565 (Beres *et al.*, 2008) in regions of the genome that are absent from *Sz*H70 and *Se*4047.

The CRISPR region of *Sz*H70 was present in 93% of *S. zooepidemicus* isolates examined by PCR, but was absent from all strains of *S. equi* tested (Figure 2.6).

Table 2.4 BLASTN analysis of *Sz*H70 spacer regions in the clustered, regularly interspaced short palindromic repeat (CRISPR) locus

Spacer	BLASTN alignment	Organism	CDS
#1 (33 nt)	Query 13 TTTTTTGACCGTCAATCAT 31 Sbjct 1768789 TTTTTTGACCGTCAATCAT 1768771 100% Identity	*Cyanothece sp.* (CP001291)	chaperone DnaJ domain protein
#2 (33 nt)	Query 1 TTTGATGTAATCTGCGATA 19 Sbjct 97504 TTTGATGTAATCTGCGATA 97486 100% Identity	*Aspergillus niger* (AM270371)	-
#3 (33 nt)	**Query 1 TTTGTTAAGGTGCTTGCTCTATTGCAAGATATC 33 Sbjct 678830 TTTGTCAAGGTGCTTGCTCTATTGCAAGATATC 678862 96% Identity**	***S. pyogenes* Manfredo (AM295007)***	**φMan.2 putative phage tail protein**
#4 (33 nt)	Query 11 TGCAAATTACAAGATTGATATTG 33 Sbjct 195 TGCAAATTACAAGATTGATATTG 217 100% Identity	*Plasmodium berghei* (XM_669489)	BIR protein
#5 (36 nt)	Query 11 TTCTCTTAGTTGGTTTACT 29 Sbjct 2151305 TTCTCTTAGTTGGTTTACT 2151287 100% Identity	*Microcystis aeruginosa* (AP009552)	hypothetical protein
#6 (34 nt)	**Query: 34 ATGTTGGTTGGACTGTTCGTCAATGTTTGATTTG 1 Sbjct: 155982 ATGTTGGTTGGACTGTTCGTCAATGTTTGATTTG 156015 100% Identity**	***Se4047***	**φSeq1 SEQ0163 phage DNA methylase**
#7 (33 nt)	**Query: 1 GTGTTCCGTTTGAAAAGCGAGTTTAGCTGCTTT 33 Sbjct: 1757760 GTGTTCCGTTTGAAAAGCGAGTTTAGCTGCTTT 1757792 100% Identity**	***Se4047***	**φSeq3 SEQ1743 phage major tail protein**
#8 (34 nt)	**Query: 1 TTTAGCCCTGTCGTTCCAACTACTTTTAAAGATA 34 Sbjct: 1758654 TTTAGCCCTGTCGTTCCAACTACTTTTAAAGATA 1758687 100% Identity**	***Se4047***	**φSeq3 SEQ1745 hypothetical phage protein**
#9 (35 nt)	Query 1 TCTGTTGGTAGGGTACAAACCCCAAC 26 Sbjct 20707 TCTGTTGGTAGAGTACAAACACCAAC 20732 92% Identity	*S. pneumoniae* AP200 (EF469826)	erythro-mycin resistance transposon Tn1806
#10 (33 nt)	Query 9 CATTACCACCAAACTCAAACA 29 Sbjct 167 CATTACCACCAAACTCAAACA 187 100% Identity caa caactgtaac caccgcggct	*Vibrio cholera* (L03220)	heat-stable enterotoxin (sto)
#11 (34 nt)	**Query 1 GTGGGGTACTTGAATGAGCAATCTAGTATTATCG 34 Sbjct 790586 GTGGGGTGCTTGAATGAGCAATCTAGTTTTATCG 790619 94% Identity**	***S. pyogenes* M1 (AE004092)***	**370.3 hypothetical phage protein**
#12 (34 nt)	Query 10 TGGATACCTTTATCCAAG 27 Sbjct 3410572 TGGATACCTTTATCCAAG 3410555 100% Identity	*Pseudomonas putida* (CP000926)	TPR repeat-containing protein
#13 (35 nt)	Query 7 GGTGGAAAAACGGCTTGC 24 Sbjct 72868 GGTGGAAAAACGGCTTGC 72851 100% Identity	*Bacteroides uniformis* (AY345595)	erythro-mycin resistance transposon CTnBST
#14 (37 nt)	Query 17 AGAATTACTTCGAGGATGG 35 Sbjct 1282296 AGAATTACTTCGAGGATGG 1282278 100% Identity	*Leishmania braziliensis* (AM494970)	-
#15 (34 nt)	**Query: 34 GAGACCACCATTGATTGAATGTATTTGATCTTTT 1 Sbjct: 1742700 GAGACCACCATTGATTGAATGTATTTGATCTTTT 1742733 100% Identity**	***Se4047***	**φSeq3 SEQ1727 *seeM***
#16 (35 nt)	Query 2 GATGAAGCTGCTACAAAGC 20 Sbjct 3830 GATGAAGCTGCTACAAAGC 3812 100% Identity	*Danio rerio* (zebrafish XM_689207)	-

Table 2.4 continued

Spacer	BLASTN alignment	Organism	CDS
#17 (35 nt)	Query 1 TTTGGCATGCCTCACATCAGACTCATGTCTAGCAT 35 |||||||| ||||||||||||||| |||||| Sbjct 1336645 TTTGGCATGTCTCACATCAGACTCATGTTTAGCAT 1336679 94% Identity	*S. pyogenes* MGAS10270 (CP000260)	putative cell surface protein
#18 (35 nt)	Query: 2 CGACACTGCGAACTAATCATTTGTGACTCCTTTC 35 ||||||||||||||||||||||||||||| Sbjct: 1745050 CGACACTGCGAACTAATCATTTGTGACTCCTTTC 1745083 100% Identity Query: 35 GAAAGGAGTCACAAATGATTAGTTCGCAGTGTCG 2 ||||||||||||||||||||||||||||| Sbjct: 177618 GAAAGGAGTCACAAATGATTAGTTCGCAGTGTCG 177651 100% Identity	*Se*4047	φSeq3 SEQ1729 putative phage amidase; φSeq1 SEQ0190 putative phage cell wall hydrolase

*Matches occur to phage proteins in additional *S. pyogenes* strains. Highlighted rows indicate full length (or near full length) spacer matches to CDSs in *S. pyogenes* (blue) and *Se*4047 (red).

2.4.1.6 Nutritional capabilities

2.4.1.6.1 Carbohydrate metabolism

Comparison of the genome sequences identified a 5 kb deletion in the *Se*4047 genome that partially deleted *lacD* and *lacG* and deleted *lacE*, *lacF* and *lacT*, which encode the sugar phosphotransferase system (PTS) proteins and associated enzymes for lactose utilisation (Appendix Table A.2). *Se*4047 also contains a deletion that removes *sorD* and alters the N-terminus of *srlM* (SEQ0286) in the sorbitol operon (Appendix Table A.2). *sorD* encodes sorbitol-6-phosphate 2-dehydrogenase required to convert D-sorbitol-6-phosphate to β-D-fructose-6-phosphate, whilst SrlM is a transcription antiterminator. The region between SEQ0536 and SEQ0537 that spans the operon required for ribose fermentation (*rbsRKDACB*) is also missing from *Se*4047 (Appendix Table A.2). All 26 *S. equi* strains examined lacked *lacE*, *sorD* and *rbsD* and the capacity to ferment lactose, sorbitol or ribose. However, only 15 (ST-7, ST-39, ST-57, ST-97 and ST-106) and one (ST-39) of 140 *S. zooepidemicus* isolates tested did not ferment ribose or sorbitol, respectively (Figure 2.6). With the exception of 2 isolates of ST-7, all other *S. zooepidemicus* isolates unable to ferment ribose also lacked the *rbsD* gene. *sorD* was present in the *S. zooepidemicus*

isolate that was unable to utilise sorbitol (ST-39), which points to different mutation in the fermentation pathway in comparison to *Se*4047. Complete lactose, sorbitol and ribose operons were present in *Sz*MGCS10565.

An 11 bp deletion in SEQ1613 (*Se*4047), which encodes a cellobiose-specific IIC PTS component, has introducted a frameshift after codon 313 and an early stop codon 18 residues later. The *Se*4047 genome also has a nonsense mutation (TAC to TAA after codon 139) that has inactivated a sorbose superfamily PTS permease gene (SEQ0871) and an ISSeq3 element has disrupted a putative β-galactosidase gene (SEQ0866) in the same operon. The sorbose EII superfamily includes sorbose-specific transporters and the broad range permeases capable of transporting mannose, fructose and N-acetylgalactosamine. Bga, the characterised β-galactosidase of *Xanthomonas manihotis* with most homology to SEQ0866 (40% amino acid sequence identity), specifically cleaves β1-3- and β1-4- linked galactose and resembles the acid β-galactosidases of eukaryotes capable of cleaving β1-3- and β1-4- linked galactose from glycoproteins and glycolipids (Taron *et al.*, 1995). These CDSs are intact in *Sz*H70 and *Sz*MGCS10565.

2.4.1.6.2 Branched-chain amino acid uptake

The *Sz*H70 and *Sz*MGCS10565 genomes encode a 50 kDa membrane protein (SZO14690/Sez_0528) with similarity to the low-affinity, branched-chain amino acid carrier of *Lactobacillus delbruckii* subsp. *lactis* BrnQ (37% amino acid sequence identity). SZO14690 (and Sez_0528), like BrnQ, contains 12 predicted transmembrane helices (TMHs). ISSeq3-mediated deletion has removed this gene from the *Se*4047 genome. SZO14690 was absent from all 26 *S. equi* isolates (Figure 2.6) but was a consistent feature of all 140 *S. zooepidemicus* isolates tested.

2.4.1.7 Hyaluronate lyase

The *Sz*H70 genome contains a single CDS encoding a putative secreted HA lyase (SZO06680). SEQ1479, the *Se*4047 orthologue, contains a 4 bp deletion (TCTC) leading to a frameshift at codon 199. An early stop codon would truncate the produced protein at amino acid 217 (including 18 aberrant residues at the C-terminus). φSeq3, a prophage of *Se*4047 encodes a different HA lyase (SEQ2045) with a primary role most likely to be involved in phage penetration through the bacterial capsule. This phage enzyme lacks a signal sequence and is therefore only likely to reach the extracellular environment following bacterial cell lysis. PCR was performed using a forward primer that covered the 4 bp of SZO06680 deleted in strain *Se*4047 to test for the presence of these 4 bp in other strains of *S. equi* and *S. zooepidemicus*. All 26 isolates of *S. equi* tested and one strain of *S. zooepidemicus* JKS115 (ST-57) were negative by PCR, whereas all other *S. zooepidemicus* isolates were positive (Figure 2.6). Sequencing revealed an IS element 905 bp from the translational start of this gene in the *S. zooepidemicus* JKS115 isolate that was negative by PCR (Holden, 2009).

2.4.1.8 ESAT-6 secretion system

A cluster of 9 genes with similarity to the 8 genes of the ESAT-6 secretion system (Ess) of *S. aureus* is present in the *Sz*H70 genome (Burts *et al.*, 2005; Burts *et al.*, 2008; Sundaramoorthy *et al.*, 2008) (Figure 2.9).

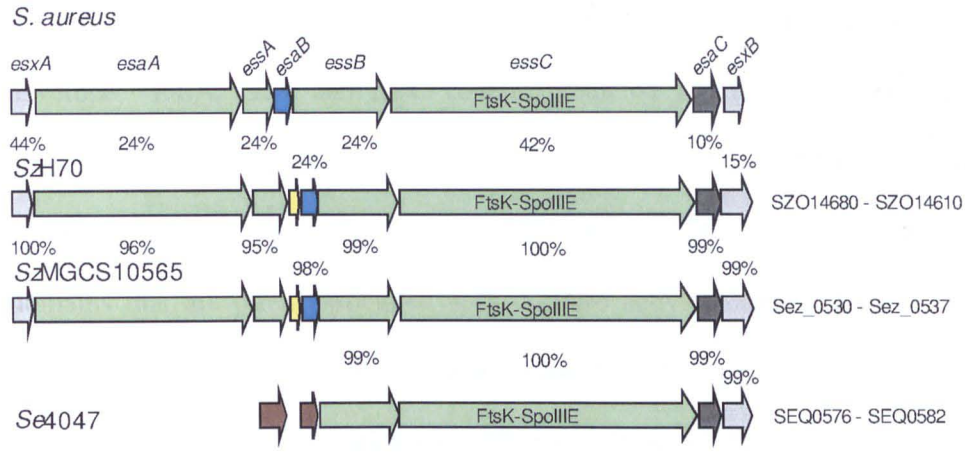


Figure 2.9 Schematic drawing of the *ess* cluster in *S. aureus*, *Sz*H70, *Sz*MGCS10565 and *Se*4047

Colouring indicates function according the the characterisation of the *S. aureus* Ess system (Burts *et al.*, 2005; Burts *et al.*, 2008; Sundaramoorthy *et al.*, 2008) and the prediction of TMHs in putative membrane proteins; light grey: ESAT-6 like secreted WXG100 proteins; dark grey: non-WXG100 secreted protein (less well conserved CDS of the *ess* clusters in Gram-positive bacteria); green: predicted membrane proteins; blue: cytoplasmic post-transcriptional regulator; brown: partial or pseudogene; yellow: possible hypothetical protein. 2 FtsK-SpoIIIE domains, a characteristic of ATPases, are present in *essC* and the *S. equi* subspecies homologues.

The extra CDS (SZO14650) in *Sz*H70 is a possible hypothetical protein of 44 amino acids. *Sz*MGCS10565 has a near identical gene cluster, whilst ISSeq3 elements have mediated the deletion of the first 2 genes in the homologous locus of *Se*4047. In addition, SEQ0576 is missing an N-terminal portion and SEQ0578 has a frameshift after codon 13 (and an early stop 13 codons later). PCR analysis showed that none of the 26 isolates of *S. equi*, but 138 of 140 *S. zooepidemicus* isolates examined contained the *esaA* gene (Figure 2.6). EsxA and EsxB are referred to as leaderless WXG100 secretion proteins due to their lack of an N-terminal signal sequence, their approximate size of 100 amino acids and the presence of a centrally located WXG-motif. These features are also conserved in the homologous proteins of the *S. equi* subspecies (where present). The role of the WXG-motif is unknown at present but may be involved in the formation of the EsxA-EsxB complex. EssA, EssB and EssC are predicted membrane proteins required for the

synthesis of EsxA and EsxB and possibly their secretion across the staphylococcal envelope. EssA, EssB and EssC contain predicted TMHs (1:1:2 respectively) and the homologues in the *S. equi* subspecies also contain predicted TMHs (2:1:1 respectively). EssC and its homologues in *Se*4047, *Sz*H70 and *Sz*MGCS10565 contain 2 FtsK/SpoIIIE domains that are often associated with ATPase activity. These proteins may therefore function as an ATPase to provide energy for the active export of secreted factors. The exact function of EsaA is unknown; although this protein has 6 predicted THMs like the homologous proteins in *S. zooepidemicus.* EsaB is a cytoplasmic protein involved in the post-transcriptional negative regulation of EsaC, and EsaC is another secretion substrate of the Ess pathway (Burts *et al.*, 2008). Genes located immediately downstream of *esxB* in *S. aureus* and the *S. equi* subspecies are not depicted in Figure 2.9. However, their conservation in other related clusters from low G + C organisms suggest that they may also be associated functionally with the Ess cluster. *Se*4047 has only one downstream gene (SEQ0583) before another IS mediated deletion event has removed a large region of DNA that includes the previously described CRISPR locus.

2.4.1.9 Factor H binding protein

Se18.9 (SEQ0235) binds to the complement regulator factor H, decreases C3b deposition and reduces the bactericidal activity of equine neutrophils towards *S. equi* (Tiwari *et al.*, 2007). This 18.7 kDa protein is encoded with 4 other genes (including 2 pseudogenes) in a region of the genome that is novel to *Se*4047 in comparison to *Sz*H70 and *Sz*MGCS10565. A low GC content suggests that this locus has been acquired through lateral gene transfer. Single CDSs with similar N-terminal signal sequences are present in the 2 *S. zooepidemicus* genomes but these only share 24 to 28% aa sequency identity with SEQ0235 overall. PCR analysis showed that all 26 isolates of *S. equi*, but only one out of 140 *S. zooepidemicus* isolates examined contained the SEQ0235 gene (Figure 2.6).

2.4.1.10 Surface proteins

2.4.1.10.1 Gene loss

Like other Gram-positive bacteria, *Se*4047 and *Sz*H70 display an array of cell wall-anchored proteins; however the genome of *Se*4047 has undergone notable gene decay in this subset of proteins in comparison to *Sz*H70 and *Sz*MGCS10565 (Table 2.5). *Sz*H70 encodes 34 intact cell wall-anchored proteins and 5 predicted pilus subunits that are likely to contribute to the formation of 2 mature pilus structures. 3 of these pilus subunits, encompassing one of the mature pilus structures, are absent from the *Se*4047 genome due to an IS-mediated deletion event (see below). The *Se*4047 genome contains homologues to the other 2 putative pilus subunits and to 33 of the 34 *Sz*H70 cell wall-anchored protein CDSs, although 8 are pseudogenes or partial genes. 2 of these *Se*4047 pseudogenes encoding Fne (fibronectin- and collagen-binding protein) and FneD (collagen-binding protein) were previously described (Lannergard, 2006; Lindmark *et al.*, 2001). The introduction of early stop codons has truncated Fne and FneD resulting in their likely secretion without a cell wall anchor and without 300 to 131 C-terminal amino acid residues, respectively. For Fne, this truncation results in the loss of a 2nd fibronectin-binding site (Lindmark *et al.*, 2001). Frameshift mutations have introduced similar early stops after codon 9 in a SeM-like surface protein (SEQ0232), after codon 182 in a SzPSe-like surface protein (SEQ0566) and after codon 58 in a putative cell wall-anchored C5A peptidase (SEQ1077). Further gene degeneration has left 3 partial genes in *Se*4047; for 2 of these (SEQ1226 and SEQ1307a) only small gene fragments remain, whilst one lacks an N-terminal signal sequence required to direct its extracellular secretion (SEQ0556). SEQ1226 is part of a bacteriocin gene cluster that has likely degraded following its disruption by insertion of an ISSeq3 element and a subsequent intra-replichore recombination event. SZO09950, the *Sz*H70 homologue of SEQ1226 contains a putative Pfam thioredoxin domain and may be involved in bacteriocin modification. Further

characterisation of SZO08560, the intact version of SEQ1307a in *Sz*H70 is presented below.

Conversely, *Se*4047 SEQ2190 encodes a surface anchored protein that has only a partial gene homologue in *Sz*H70 (SZO18970) and an unconfirmed pseudogene homologue in *Sz*MGCS10565 (Sez_1908-1910) suggesting that gene degradation has occurred in *Sz*H70 and *Sz*MGCS10565 subsequent to their divergence from the ancestral *S. zooepidemicus* strain from which *S. equi* evolved. Interestingly, all diverse strains of *S. zooepidemicus* tested for the presence of an N-terminal region of SEQ2190 were negative by PCR (Figure 2.6), which implies a lack of SEQ2190 homologues (at least with the same nucleotide sequence in the primer binding site(s)) in the general *S. zooepidemicus* population. SEQ2190 shares amino acid sequence identity (30%) with the C-terminal half of a GAS M3 protein that binds multiple ligands including fibrinogen, fibronectin and serum albumin (Reichardt *et al.*, 1995), but has a unique N terminal region in comparison with other proteins in the public databases.

Table 2.5 Sortase-processed surface proteins of *Se*4047, *Sz*H70 and *Sz*MGCS10565

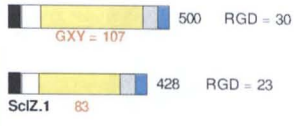
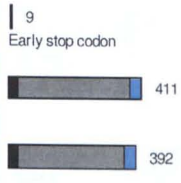
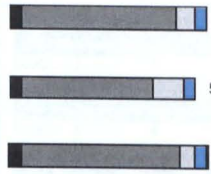
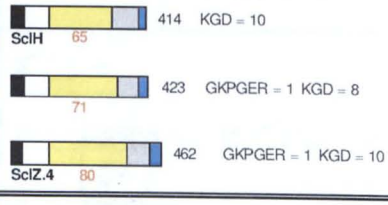
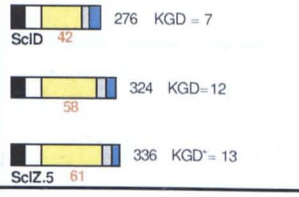
CDS	% ID vs SEQ	% ID vs SZO	% ID V vs SEQ	% ID V vs SZO	Product (putative sortase motif)	Schematic of protein homologues (see below for key)
NP SZO00830 Sez_0099	- -	 85.2	- -	 100	putative collagen-like surface-anchored protein (LPATG) Nearly identical except for variation in the number of GXYs	
SEQ0090 SZO00840 Sez_0100	 52.6 51.1	 79.4	 15.6 13.9	 65.6	putative collagen-like surface-anchored protein SclG (LPATG)	
NP NP Sez_0162	 - -	 -	 - -	 -	putative collagen-like surface-anchored protein (LPATG)	
SEQ0232† SZO01430 Sez_0164	 - -	 64.7	 - -	 -	SeM-like cell surface-anchored protein (LPSTG)	
SEQ0256 SZO01590 Sez_0185	 74.3 85.4	 79.0	 - -	 -	putative cell surface-anchored protein (LPATA)	
SEQ0260 SZO01630 Sez_0190	 82.2 83.3	 88.5	 96.2 96.2	 100	putative collagen-like cell surface-anchored protein SclH (LPATG)	
SEQ0280 SZO17540 Sez_0213	 85.2 82.1	 96.4	 100 100	 100	putative collagen-like cell surface-anchored protein SclD (LPATG) 100% identical except for variation in the number of GXYs	

Table 2.5 continued

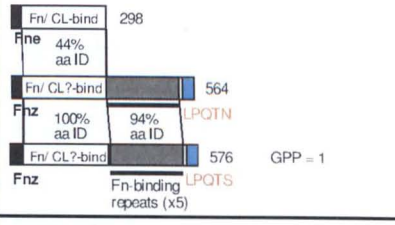
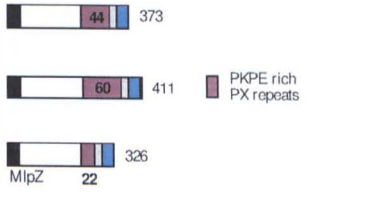
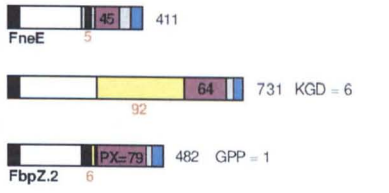
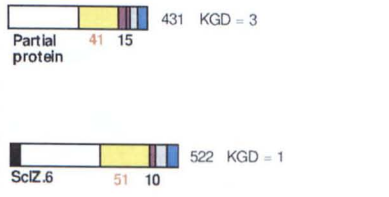
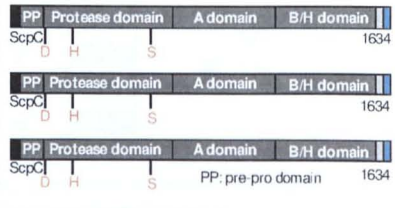
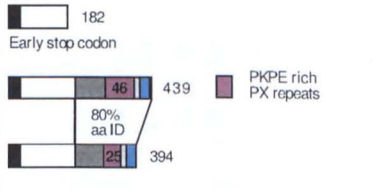
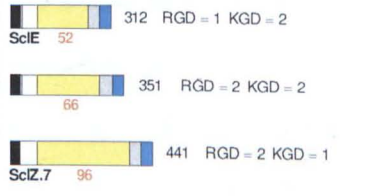
CDS	% ID vs SEQ	% ID vs SZO	% ID V vs SEQ	% ID V vs SZO	Product (putative sortase motif)	
SEQ0375†					fibronectin and collagen binding protein Fne/Fnz (LPQTN/S)	
SZO16630	-		43.6			
Sez_0313	-	65.8	43.6	100		
SEQ0402					putative cell surface-anchored protein (LPSTG)	
SZO16370	61.6		41.4			
Sez_0338	52.2	53.2	40.3	47.5		
SEQ0555					collagen and fibronectin-binding cell surface-anchored protein FneE (LPRTN)	
SZO14890	39.0		45.7			
Sez_0499	63.7	42.9	62.6	46.0		
SEQ0556†					putative collagen-like surface-anchored protein (LPKTN)	
NP	-		-			
Sez_0500	-	-	-	-		
SEQ0563					chemokine protease ScpC (LPSTG)	
SZO14810	97.8		-			
Sez_0511	98.6	98.2	-	-		
SEQ0566†					SzP/SzPSe-like cell surface-anchored protein (LPSTG)	
SZO14790	-		27.1			
Sez_0513	-	67.0	31.2	43.4		
SEQ0633					putative collagen-like surface-anchored protein SclE (LPATG)	
SZO13850	75.0		45.1			
Sez_0608	61.8	78.5	45.4	98.0		

Table 2.5 continued

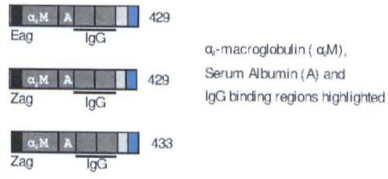
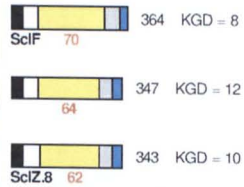
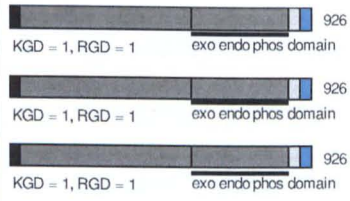
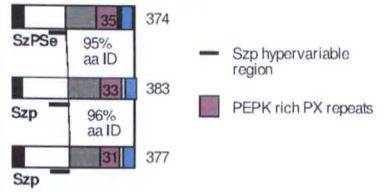
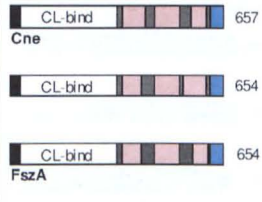
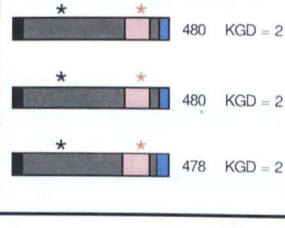
CDS	% ID vs SEQ	% ID vs SZO	% ID V vs SEQ	% ID V vs SZO	Product (putative sortase motif)	
SEQ0646					putative cell surface-anchored protein (LPLTG)	
SZO13730	95.5		-			
Sez_0623	93.4	92.8	-	-		
SEQ0721					Ig, α_2M and albumin binding protein Eag/Zag (LPTTG)	
SZO13070	99.5		-			
Sez_0696	95.4	95.4	-	-		
SEQ0855					putative collagen-like cell surface-anchored protein SclF (LPSTG)	
SZO12230	77.0		59.6			
Sez_0729	77.2	90.2	50.0	52.0		
SEQ0904					endonuclease/ exonuclease/ phosphatase (exo endo phos) family surface anchored protein (LPKTG)	
SZO11790	97.4		-			
Sez_0781	95.0	95.9	-	-		
SEQ0933					SzPSe: fibrinogen-binding cell surface-anchored protein; SzP: adhesin and hypervariable protective antigen (LPSTG)	
SZO11530	79.3		54.5			
Sez_0809	78.2	90.6	53.2	80.4		
SEQ0935					collagen binding, putative ancillary pilus subunit Cne (LPDTG)	
SZO11510	88.9		88.2			
Sez_0811	87.1	94.2	86.5	91.2		
SEQ0936					putative backbone pilus subunit (T6-antigen-like) (LPSTG)	
SZO11500	98.3		-			
Sez_0812	96.7	95.0	-	-		

Table 2.5 continued

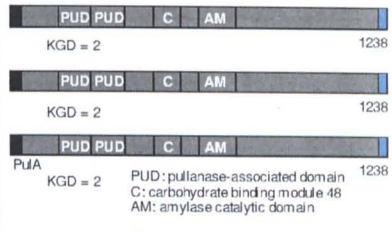
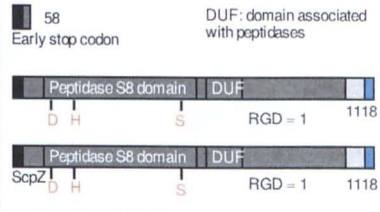
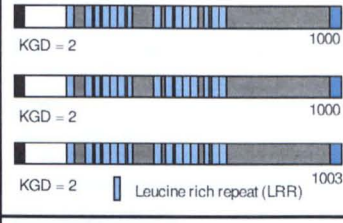
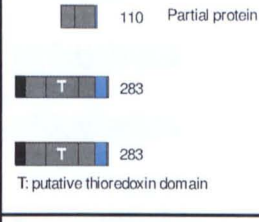
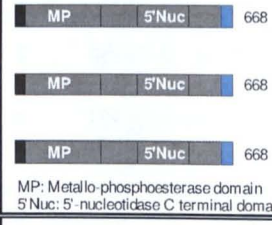
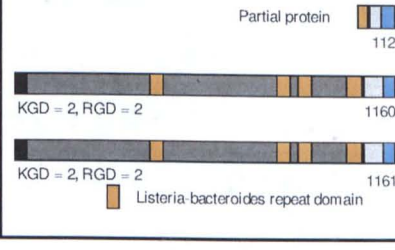
CDS	% ID vs SEQ	% ID vs SZO	% ID V vs SEQ	% ID V vs SZO	Product (putative sortase motif)	
SEQ0939					putative cell surface-anchored protein (LPSTG)	
SZO11470	66.7		57.9			
Sez_0815	73.9	76.6	38.0	48.5		
SEQ0944					putative cell surface-anchored pullulanase (LPKTG)	
SZO11410	98.2		-			
Sez_0820	97.9	99.0	-	-		
SEQ1077†					putative cell surface-anchored C5A peptidase (LPKTS)	
SZO10150	-		-			
Sez_0946	-	97.6	-	-		
SEQ1116					putative adhesin (LPKTG) (LrrG-like, *S. agalactiae*)	
SZO08940	93.2		84.2			
Sez_1073	94.5	94.4	87.2	88.8		
SEQ1226†					putative cell surface-anchored thioredoxin (FPKTG) (located into bacteriocin cluster)	
SZO09950	-		-			
Sez_0970	-	100	-	-		
SEQ1278					putative cell surface-anchored 5'-nucleotidase (LPATG)	
SZO08800	99.4		-			
Sez_1088	99.1	99.1	-	-		
SEQ1307a†					Listeria-Bacteroides repeat domain containing cell surfaced-anchored protein (LPKTG)	
SZO08560	-		-			
Sez_1114	-	90.4	-	-		

Table 2.5 continued

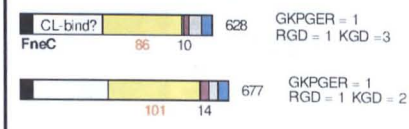
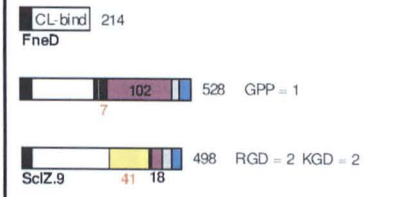
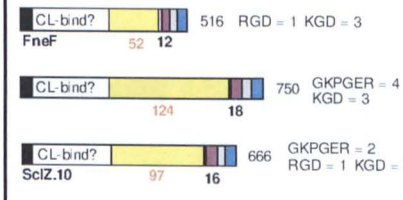
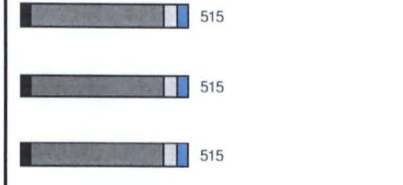
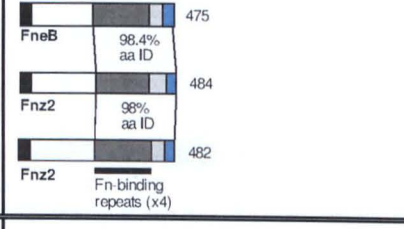
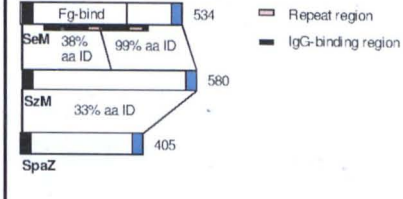
CDS	% ID vs SEQ	% ID vs SZO	% ID V vs SEQ	% ID V vs SZO	Product (putative sortase motif)	
SEQ1606					collagen-binding collagen-like cell surface-anchored protein FneC (LPKTN)	
SZO05380	56.4		38.5			
NP	-	-	-	-		
SEQ1607†					collagen-binding surface-anchored protein FneD (LPKTN)	
SZO05350	-		39.4			
Sez_1421	-	46.6	33.6	50.2		
SEQ1649					collagen-binding collagen-like surface-anchored protein FneF (LPKTN)	
SZO04910	59.7		88.8			
Sez_1457	53.1	67.7	53.9	54.7		
SEQ1817					putative collagen-like surface-anchored protein SclI (LPATG)	
SZO03720	74.2		19.7			
Sez_1598†	-	-	16.7	35.4		
SEQ1959					putative cell surface-anchored protein (LPRSG)	
SZO02400	95.1		-			
Sez_ 1737	96.9	97.5	-	-		
SEQ1999					fibronectin- and collagen-binding protein FneB/Fnz2 (LPKTH)	
SZO02080	76.5		45.5			
Sez_1759	76.7	97.9	45.5	99.5		
SEQ2017					antiphagocytic cell surface-anchored fibrinogen- and IgG Fc-binding protein SeM (LPSTG)	
SZO01900	65.8		-			
Sez_1778	36.6	33.1	-	-		

Table 2.5 continued

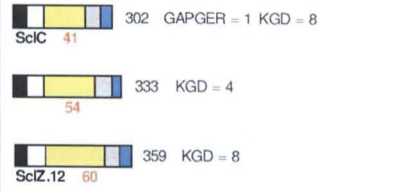
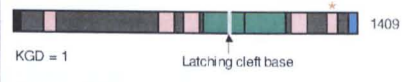
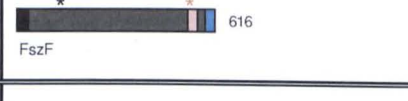
CDS	% ID vs SEQ	% ID vs SZO	% ID V vs SEQ	% ID V vs SZO	Product (putative sortase motif)	
SEQ2100 SZO18100 Sez_1802	 98.7 99.3	 98.4	 97.9 98.6	 97.9	putative cell surface-anchored protein (LPATG)	
SEQ2101 SZO18110 Sez_1803	 58.8 83.3	 65.9	 16.9 100	 16.9	putative collagen-like surface-anchored protein SclC (LPATG)	
NP SZO18330 NP	 - -	 -	 - -	 -	putative minor pilus subunit (FPMTG)	
NP SZO18320 NP	 - -	 -	 - -	 -	putative major pilus subunit (RrgB-like) (IPQTG)	
NP SZO18310 NP	 - -	 -	 - -	 -	putative minor pilus subunit (RrgC-like) (VPYTG)	
NP NP Sez_1829	 - -	 -	 - -	 -	putative major pilus subunit (T6-antigen-like) (IPNTG)	
NP NP Sez_1828	 - -	 -	 - -	 -	putative minor pilus subunit (IPYTG)	

Table 2.5 continued

CDS	% ID vs SEQ	% ID vs SZO	% ID V vs SEQ	% ID V vs SZO	Product (putative sortase motif)	
NP					putative minor pilus subunit (LPATG)	
NP	-		-			KGD = 1 * 832 FbpZ
Sez_1825	-	-	-	-		
NP					putative minor pilus subunit (LPSSG)	
NP	-		-			KGD = 1 * 967 FszD
Sez_1822	-	-	-	-		
NP					putative major pilus subunit (T6-antigen-like) (LPSTG)	
NP	-		-			RGD = 1 549 FszC
Sez_1821	-	-	-	-		
SEQ2180					putative cell surface-anchored protein (LPATG) (C-terminal region homology to GAS M/Mrp proteins)	626 99% aa ID 91% aa ID 629 46% aa ID 89% aa ID 626 CspZ.2
SZO18890	94.3		98.3			
Sez_1901	67.9	70.4	35.8	37.9		
SEQ2190					putative cell surface-anchored protein (LPATG) (C-terminal region homology to GAS M proteins)	673 82% aa ID Partial protein 372 Early stop 271 Possible frameshift
SZO18970†	-		-			
Sez_1908-1910†	-	-	-	-		

Needleman-Wunsch global alignment was used to calculate % amino acid identity of intact CDS and V regions relative to the homologues in *Se*4047 (SEQ) and *Sz*H70 (SZO) respectively. NP Not present. † Pseudogene or gene remnant. Integrin binding motifs shown (GXPGER, KGD, RGD, GPP).

Key:-

- Secretion signal peptide
- Variable region
- Collagen-like region (no. of GXY-motifs shown in red)
- PX repeats (no. of repeats shown in black)
- Proline-rich cell wall spanning region
- Cell wall anchor
- Pfam Cna_B domain (PF05738)
- SdrG IgG-like domain

* Pilin motif
* E box
Fn bind: fibronectin binding region
CL bind: collagen binding region
Fg bind: fibrinogen binding region

The *Sz*H70 and *Sz*MGCS10565 genomes encode a 131 kDa putative surface protein containing 1,160 amino acids with an LPXTG motif (SZO08560 and Sez_1114). However, the *Se*4047 genome encodes only the final 112 amino acids of this protein (SEQ1307a) and lacks an adjacent gene predicted to encode a recombinase (SZO08550 and Sez_1116) (Table 2.5). SZO08560 and Sez_1114 share sequence similarity with hypothetical proteins of *S. suis* strain 05ZYH33 (SSU05_0473) and *S. agalactiae* strain COH1 (SAN_1519) and contain 4 Listeria-Bacteroides repeat Pfam domains (PF09479). These ~70 amino acid residue repeats occur in a range of Gram-positive surface proteins including families of internalins of *Listeria* species (Figure 2.10).

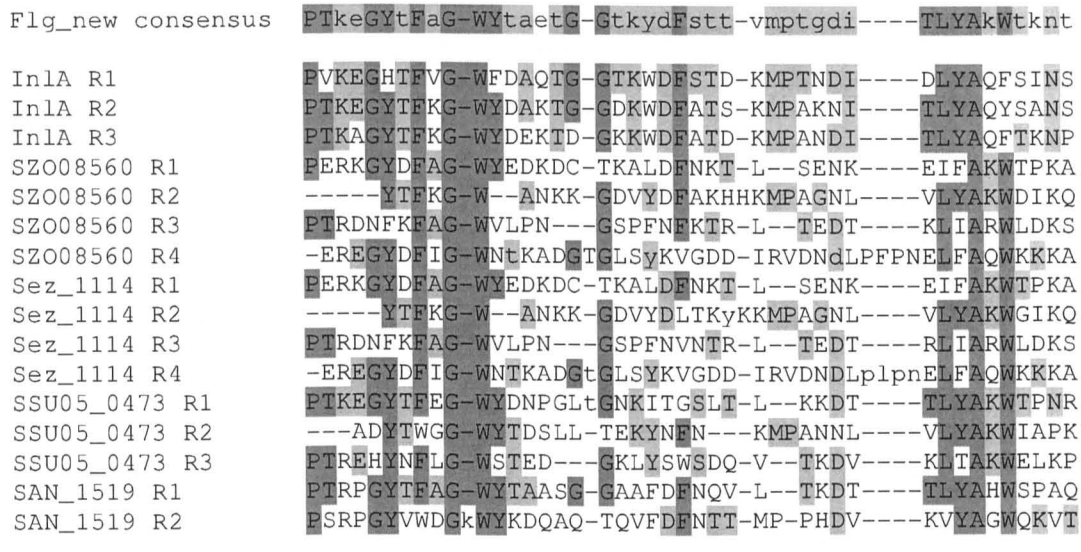


Figure 2.10 Alignment of domains in SZO08560 and similar proteins to the Pfam hidden Markov model (HMM) for the *Listeria-Bacteroides* repeat domain (Flg_new)

Flg_new repeats are a feature of some *Bacteroides forsythus* proteins and families of internalins of *Listeria* species. Matches to the highly conserved and less well conserved Flg_new residues are shown in dark and light grey respectively. 2 to 4 Flg_new repeats (R1 to R4) are shown in InlA, (*Listeria monocytogenes,* ABO32414), SZO08560 (*Sz*H70), Sez_1114 (*Sz*MGCS10565), SSU05_0473 (*Streptococcus suis* strain 05ZYH33, A4VTK0) and SAN_1519 (*Streptococcus agalactiae* strain COH1, Q3D8T2).

Examination of the *Sz*H70 genome sequencing data revealed five sequence reads that positioned the promoter region of SZO08560 (-170 bp to -55 bp) in the reverse orientation (Matt Holden, Sanger Institute). This sequence is bordered by GTAGACTTTA and TAAAGTCTAC inverted repeats; inversion of this sequence may switch transcription of SZO08560 on or off, thereby modulating the production of this surface protein (Figure 2.11).

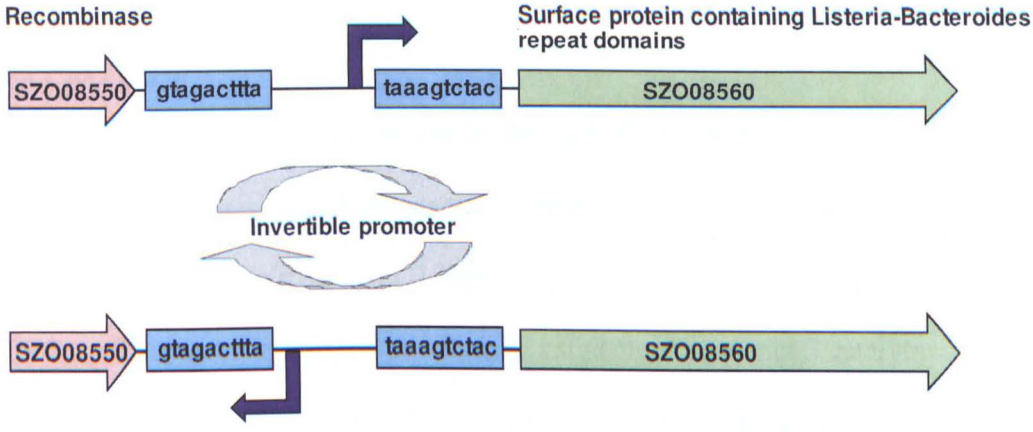


Figure 2.11 Diagram of the SZO08560 invertible promoter in *Sz*H70

The promoter region of SZO08560 (-170 bp to -55 bp), bordered by GTAGACTTTA and TAAAGTCTAC inverted repeats, inverts to switch transcription from the forward to reverse orientation.

Reverse transcription QPCR performed by Karen Steward (AHT) using RNA extracted from log-phase cultures of *Sz*H70 and normalised for expression of the housekeeping gene *gyrA* demonstrated that the SZO08560 promoter of *Sz*H70 transcribed 44-fold more RNA in the forward direction than the reverse. None of the 26 isolates of *S. equi*, but 101 of 140 *S. zooepidemicus* isolates tested positive for SZO08560 by PCR (Figure 2.6), which represents a significant difference in the prevalence of this gene between the 2 subspecies ($P < 0.0001$, Fischer's exact test). In addition, a higher proportion of *S. zooepidemicus* respiratory isolates (including those from LN abscesses) and *S. zooepidemicus* isolates collected from cases of abortion or uterine infection had the SZO08560 gene (77%, n =

118) compared to other *S. zooepidemicus* isolates (66%, n = 20, P = 0.002, Fischer's exact test). Of the *S. zooepidemicus* respiratory isolates, a significantly higher percentage of those collected from the lower respiratory tract of horses and dogs (including isolates from cases of pneumonia) had the SZO08560 gene (100%, n = 14) than those collected from the upper respiratory tract (including from LN abscesses, 71%, n = 59) (P = 0.03, Fischer's exact test). *Sz*MGCS10565 contains an IS element between the inverted repeats bordering the Sez_1114 promoter and the recombinase (Sez_1116).

2.4.1.10.1.2 Pili

2 putative sortase CDSs are present in the *Se*4047 genome: *srtA* (SEQ1171) and *srtC.1* (SEQ0937), whereas the *Sz*H70 genome contains 5: *srtA* (SZO09440), *srtC.1* (SZO11490), *srtC.2* (SZO18290), *srtC.3* (SZO18280) and *srtC.4* (SZO18270). These putative sortase CDSs all contain a conserved TLXTC signature sequence, catalytic site residues corresponding to C_{184}, H_{120} and R_{197} of SrtA in *S. aureus* and a predicted N-terminal signal sequence (Dramsi *et al.*, 2005). Those belonging to the SrtC-family (labelled *srtC.1-4*) contain features characteristic of this sortase class including a C-terminal hydrophobic domain and a conserved proline residue located after the catalytic site TLXTC (Dramsi *et al.*, 2005). As SrtC enzymes are often clustered in the genome with their cognate pilin substrates, these *srt* CDSs are predicted to be involved in pilus expression (Dramsi *et al.*, 2005).

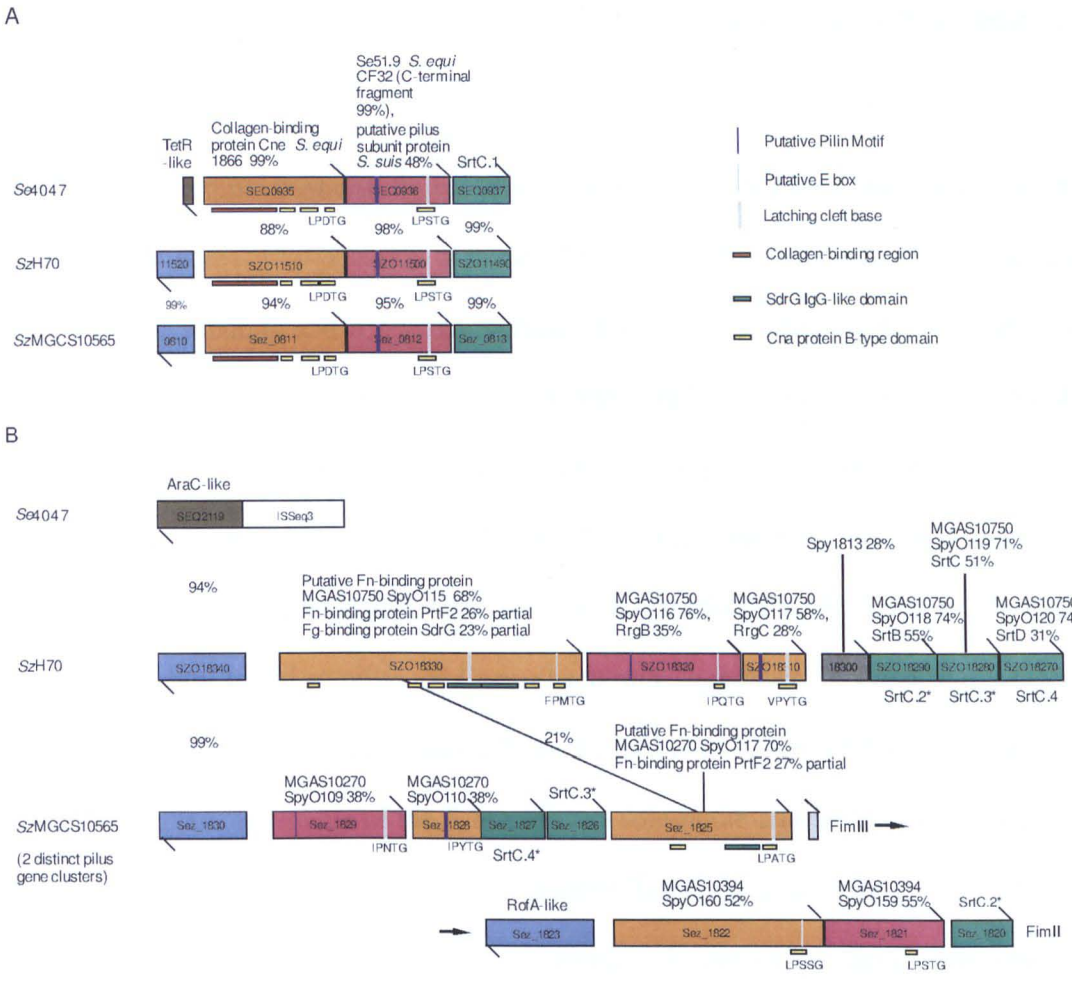


Figure 2.12 A schematic of 2 pilus loci (A and B) in *Se*4047, *Sz*H70 and *Sz*MGCS10565

Diagram shows CDSs for putative backbone/major subunits (red), putative ancillary/minor subunits (orange), C-family sortases (green) and regulatory proteins (blue). Novel CDSs encode a putative exported protein (dark grey) and a hypothetical protein (light grey). Truncated regulatory pseudogenes (brown) are present in *Se*4047 in locus I (A) and locus II (B) resulting from a single point mutation (TAC to TAA) or insertion of an IS element respectively. Recombination between ISSeq3 elements has removed the 2nd pilus locus from *Se*4047, with the exception of the truncated remnant of SEQ2119. BLASTP matches are shown above with the percentage amino acid sequence identity indicated. Needleman-Wunsch global alignments were used to conduct pairwise comparisons of CDSs in the 3 *S. equi* subspecies (percentage amino acid sequence identity shown). MGAS10750 Spy0115-Spy0120 are CDSs of a predicted pilus locus in *S. pyogenes* M4 (Beres *et al.*, 2006), SdrG is a fibriongen-binding protein of *Staphylococcus epidermidis* (Ponnuraj *et al.*, 2003), PrtF2 is a fibronectin-binding protein of *S. pyogenes* (Ramachandran *et al.*, 2004), Spy1813 encodes a hypothetical phage protein of *S. pyogenes* MGAS6180 (Green *et al.*, 2005), whilst RrgB, RrgC, SrtB, SrtC and SrtD are variants of the RlrA pilin proteins and associated sortases of *S. pneumoniae* (Moschioni *et al.*, 2008). MGAS10270 Spy0109-Spy0110 (Beres *et al.*, 2006) and MGAS10394 Spy0160-Spy0159 (Banks *et al.*, 2004) are predicted fimbrial structural subunits in *S. pyogenes* M2 and M6 respectively. In addition, Sez_1821 and Sez_1829 have homology to GAS T-antigens (major fimbrial subunit) (Beres *et al.*, 2008). Note the 2nd pilus locus is in the same genome location as the FCT locus of *S. pyogenes*, flanked by conserved CDSs encoding a heat shock protein 33 homologue and a putative toxic anion resistance protein in all 3 *S. equi* subspecies (Bessen

& Kalia, 2002). * These encoded SrtC-family proteins share 36% to 50% amino acid sequence identity.

The *Sz*H70 genome contains 2 loci that contain genes required putatively for pilus expression (Figure 2.12). The first of these (SZO11520-SZO11490) encodes a regulator, 2 putative pilin subunits and a putative sortase (SrtC.1). Orthologous loci are present in *Se*4047 (SEQ0934-SEQ0937) and *Sz*MGCS10565 (FimI: Sez_0810-Sez_0813), and share 88-99% and 94-99% amino acid sequence identity respectively with the *Sz*H70 locus. However, the TetR-like regulator SEQ0934 of *Se*4047 contains a nonsense mutation at codon 43. The predicted major pilin subunit (SZO11500/SEQ0935/Sez_0812) has the conserved E box found in pilins of *Corynebacterium diphtheria, Actinomyces naeslundii* and some streptococci (Scott & Zahner, 2006; Telford *et al.*, 2006) (Table 2.6). The highly conserved glutamate residue of the E box plays a role in the attachment of minor pilins to the shaft protein through an unknown mechanism (Ton-That *et al.*, 2004). A possible pilin motif in SZO11500/SEQ0935/Sez_0812 contains only the conserved proline and lysine residues; the latter is thought to form a peptide bond with the threonine at the C-terminus of the next subunit (LPXTG motif) allowing the oligomerization of the backbone pilin (Kang *et al.*, 2007; Scott & Zahner, 2006; Ton-That & Schneewind, 2003) (Table 2.6). The second pilus subunit (SZO11510/SEQ0935/Sez_0811) lacks an E box or pilin motif and has been previously characterised as a collagen-binding adhesin Cne of *S. equi* (Lannergard *et al.*, 2003), similar to Cna of *S. aureus* (Patti *et al.*, 1992). Cne is therefore likely to be an accessory pilin subunit.

Table 2.6 Conserved motifs in putative and proven pilin subunits

Strain(s)	Pilus locus (subunit)	Name	Pilin motif	E box	CWSS*
*Sz*H70/*Se*4047/SzM GCS10565	1 (**major**)	SZO11500/SEQ0936/Sez_0812	vvkntsfk**PK**	Yk**L**s**E**T**A**PgGY	**LPsTG**
*Sz*H70/*Se*4047/SzM GCS10565	1 (minor)	SZO11510/SEQ0935/Sez_0811	-	-	**LPdTG**
*Sz*H70	2 (minor)	SZO18310	dadvtkq**YPK**	Yy**L**k**E**v**A**PxGY	vPyTG
*Sz*H70	2 (major)	SZO18320	stikel**VYPK**	Yy**L**e**E**lt**A**PiGf	iPqTG
*Sz*H70	2 (minor)	SZO18330	-	Yaik**E**ek**A**PdGY	fPmTG
*Sz*MGCS10565	2I (major)	Sez_1829	tvedannsPK	Yy**L**k**E**v**A**PaGY	iPnTG
*Sz*MGCS10565	2I (minor)	Sez_1828	qkepllVYPK	-	iPyTG
*Sz*MGCS10565	2 (minor)	Sez_1825	-	Ye**L**w**E**ik**A**PnGY	LPaTG
*Sz*MGCS10565	2II (minor)	Sez_1822	-	Yy**L**y**E**Tk**A**raGY	LPsSG
*Sz*MGCS10565	2II (major)	Sez_1821	-	-	**LPsTG**
S. pyogenes MAS10394	(major)	T6	-	-	**LPsTG**
S. pyogenes SF370	(major)	Spy0128	-	-	evpTG
C. diptheria NCTC13129	(major)	SpaA	WLQDVHVYPK	FCLVETATASGY	LPlTG
C. diptheria NCTC13129	(major)	SpaD	WNYNVVAYPK	FCLKETKAPAGY	LPmTG
S. agalactiae COH1	(major)	GBS80	LSEDKVIYPK	YVLKEIETQSGY	iPkTG
S. pneumonia TIGR4	(major)	RrgB	DVVDAHVYPK	YYLEETKQPAGY	iPqTG
A. naeslundi T14V, type 1	(major)	FimP	WNYNVHVYPK	YCLVETKAPEGY	LPlTG
	Consensus†		WxxxVxVYPK	YxLxETxAPxGY	LPxTG

Table adapted from (Scott & Zahner, 2006). * CWSS, cell wall sorting signal motif preceding the hydrophobic stretch and charged tail. † Consensus sequence.

The second *Sz*H70 pilus cluster consists of CDSs encoding 3 putative surface proteins (SZO18310-SZO18330), one putative exported protein (SZO18300), 3 putative sortase enzymes, SrtC.2, SrtC.3 and SrtC.4 (SZO18290-SZO18270), and an AraC-like transcriptional regulator (SZO18340). A similar locus that shares 58% to 76% amino acid sequence identity to these *Sz*H70 CDSs is present in *S. pyogenes* MGAS10750 (Spy0115-Spy0120) (Beres *et al.*, 2006), although a homologue of SZO18300 is absent and the GAS locus appears to be under the control of a different RofA-like regulator (Spy0113). The GAS locus also encodes an additional putative fibronectin binding protein (Spy0114). The

genome of strain *Se*4047 lacks this putative pilus locus through ISSeq3-mediated deletion. None of the 26 isolates of *S. equi*, but 81 of 140 *S. zooepidemicus* isolates tested positive for srtC.3 or srtC.4 by PCR. This difference in prevalence between the 2 subspecies was statistically significant using Fischer's exact test ($P < 0.0001$). The genome of *Sz*MGCS10565 does not contain a homologue of this *Sz*H70 pilus locus, but instead contains two other consecutive pilus loci Fim II and Fim III at the same genome location (Beres *et al.*, 2008). Fim III is flanked by an AraC-like regulator (Sez_1830), which is orthologous to SZO18340 of *Sz*H70.

Further analysis was performed in order to probe the binding activity of this *Sz*H70-specific pilus. SZO18330 had most overall homology to Spy0115 (putative fibronectin binding protein) (Beres *et al.*, 2006) and partial homology with SdrG (Ponnuraj *et al.*, 2003) and PrtF2 (Ramachandran *et al.*, 2004) (23 – 26% amino acid sequence identity), which bind fibrinogen (*Staphylococcus epidermidis*) and fibronectin (*S. pyogenes*), respectively. The homologous regions in SdrG include tandem IgG-like domains that allow folding of the protein to create a docking groove for fibrinogen binding between these domains (Ponnuraj *et al.*, 2003). In addition, a conserved sequence motif TYTFTDYVD of SdrG is also present in one of the IgG-like folded domains of SZO18330 (IYTFTDYVE), and represents the base of a potential latching cleft into which a latching strand could insert to stabilize an adhesin-ligand complex. This motif and IgG-like folds are a feature of Cna and a large family of structurally related putative Gram-positive MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) that are likely to use a similar 'dock, lock and latch' model to bind short linear peptide ligands (Ponnuraj *et al.*, 2003). Residues that interact with the fibrinogen peptide in SdrG are not conserved in SZO18330 suggesting that a different linear peptide is accommodated by this encoded protein in *Sz*H70. Cna protein B-type domains are also a feature of SZO18330, SdrG and other related putative adhesins. In Cna, these are thought to form a stalk to

present the ligand binding domain away from the bacterial cell surface. PrtF2 has similar IgG-like folds and Cna protein B-type domains that align to SZO18330 but the 4 Pfam fibronectin binding repeat domains of PrtF2 that utilise a tandom β-zipper mechanism for fibronectin binding (Rakonjac *et al.*, 1995; Schwarz-Linek *et al.*, 2003) are not present in SZO18330. SZO18330 is a predicted ancillary pilin subunit and doesn't contain a pilin motif, although an E box is present suggesting that this isn't just a feature of backbone pilin subunits.

SZO18320 and SZO18310 share 35% and 28% amino acid sequence identity with RrgB and RrgC, the major and ancillary pilins respectively encoded from the *rlrA* genetic islet of *S. pneumoniae* (Moschioni *et al.*, 2008) (Figure 2.12). The shaft protein of this putative pilus in *Sz*H70 is therefore predicted to be encoded by SZO18320, whilst SZO18310 is likely to encode a minor subunit of the pilus, although surprisingly a putative pilin motif and E box is present in both proteins (Table 2.6). E box-like motifs have been identified in the accessory subunits of some GBS pilins (Rosini *et al.*, 2006) and the *S. pneumoniae* RlrA pilus, which in the latter are able to form heterodimers (Falker *et al.*, 2008; LeMieux *et al.*, 2008). SrtC.2, SrtC.3 and SrtC.4 also have homology to the 3 sortases of the *rlrA* cluster SrtB, SrtC and SrtD (55%, 51% and 31% amino acid sequence identity respectively) (Moschioni *et al.*, 2008), which have recently been renamed SrtC.1, SrtC.2 and SrtC.3 (LeMieux *et al.*, 2008). Interestingly, the *Sz*H70 SrtC.4 is more divergent in sequence from the other 2 Srt enzymes and like SrtD (SrtC.3) of the RlrA locus might be hypothesised to play a similar role in the non-homogenous focal presentation of the pilus on the bacterial cell surface (Falker *et al.*, 2008).

All 3 subunit proteins have divergent CWSS motifs with a variant amino acid (phenylalanine, isoleucine or valine) replacing the canonical leucine in position one (Table 2.6, Figure 2.12). Such variation in the CWSS has been shown at least in part to contribute

to sortase specificity, in particular for the recognition and incorporation of individual accessory subunits into the mature fimbrial structure (Barnett *et al.*, 2004; Gaspar & Ton-That, 2006; LeMieux *et al.*, 2008).

2.4.1.10.2 Genetic diversification in surface proteins

Substantial amino acid sequence variation is evident in many of the *Se*4047 surface proteins when compared to their homologues in *Sz*H70 (and *Sz*MGCS10565) (Table 2.5). 13 *Se*4047 CDSs share only 40 to 79% amino acid sequence identity with their *Sz*H70 homologues and, with the exception of FneF, display even greater variation in their N-terminal regions (16 to 60% amino acid sequence identity). 8 of these encode collagen-like proteins and are discussed in more detail below. The other 5 include 2 surface proteins of unknown function containing C-terminal proline-rich repeats (SEQ0402 and SEQ0939), a fibrinogen-binding protein SzPSe (Timoney *et al.*, 1997b), the antiphagocytic fibrinogen and IgG-binding protein SeM (Meehan *et al.*, 2002; Timoney *et al.*, 1997b) and a fibronectin-binding protein FneB (Lannergard *et al.*, 2005). Generally, the function of equivalent proteins in *Sz*H70 has not been determined, although Szp, the SzPSe-homologue, exhibits hypervariability amongst *S. zooepidemicus* isolates and functions as an adhesin in the ATCC 35246 strain of *S. zooepidemicus* (Fan *et al.*, 2008) (86% amino acid sequence identity to Szp *Sz*H70). Fnz2 (*S. zooepidemicus* strain VTU211), like its homologue FneB (*Se*4047) binds fibronectin via a C-terminal repeat region (Hong, 2005; Lannergard, 2006). Fnz2, but not FneB, is also able to bind collagen which may be explained through the differences that occur in the N-terminal domain of these proteins. A near identical homologue of Fnz2 (strain VTU211) is present in *Sz*H70 and *Sz*MGCG10565.

In addition, the 3 *Se*4047 pseudogenes likely to encode secreted proteins lacking a cell wall anchor, Fne, SEQ0566 (SzPSe-like) and FneD, share only 27 to 44% amino acid sequence

identity with the N-terminal regions of their *Sz*H70 homologues. The protein homologues often differ in size, which in many cases is influenced by the length of proline-rich C-terminal regions and the number of proline-rich repeats. Such domains may aid in the presentation of the ligand-binding site away from the bacterial cell surface.

2.4.1.10.2.1 Collagen-like surface proteins

Like other cell wall-anchored proteins, Scls contain an N-terminal signal sequence, followed by a unique non-repetitive region (V region), a proline-rich putative wall spanning region (W) preceding an LPXTG-motif and a hydrophobic transmembrane region (cell wall anchor). These proteins also contain contiguous tracks of Glycine-Xxx-Yyy (GXY) repeat motifs between the V and W domains, which confer an alfa-helix secondary structure (with the glycine acting as a backbone) similar to that of mammalian collagen (Han *et al.*, 2006b; Xu *et al.*, 2002). 11 *scls* are encoded in the genome of *Sz*H70 (Table 2.5). In comparison, *Se*4047 has 9 corresponding Scl genes but lacks a homologue of the SZO00830 encoded Scl or a collagen-like region in FneE (Table 2.5). FneE, a collagen- and fibronectin-binding surface protein of *S. equi* (Lannergard, 2006), contains a string of 2 + 3 GXY-motifs interrupted by 4 amino acids, which is a possible remnant of a collagen-like domain. Similar interrupted GXY-motifs are present in Sez_0499, the FneE homologue in *Sz*MGCS10565, whereas the counterpart in *Sz*H70 (SZO14890) has a long collagen-like domain containing 92 GXY sequences. SZO05350, a putative collagen-binding surface protein of *Sz*H70, also contains possible traces of an ancestral collagen-like domain. GXY-motifs are present in SclZ.9 (n = 41) but not FneD, the respective SZO05350 homologues in *Sz*MGCS10565 and *Se*4047. The genome of *Sz*MGCS10565 contains 11 *scls* (and an unconfirmed Scl pseudogene), with a novel CDS (*sclZ.3*) relative to *Se*4047/*Sz*H70 downstream of the transcription antitermination protein *nusG* gene. Another Scl gene *sclZ.6* is unique to this epidemic nephritis strain of *S. zooepidemicus* in

comparison to *Sz*H70 although a partial gene homologue, lacking a secretion signal sequence is present in *Se*4047 (SEQ0556).

FneE and 7 Scl CDSs in *Se*4047 (SclG, FneE, FneC, SclC, SclE, SclF, SclI, FneF) show considerable sequence divergence when compared to their Scl homologues in *Sz*H70 (Table 2.5). In 5 (SclG, FneE, FneC, SclC, SclE), the variation occurs in both the V and collagen-like regions with the level of amino acid sequence identity shared between the *Se*4047 and *Sz*H70 homologues ranging from 16% to 46% in the V region and 4% to 63% in the collagen-like region. SclF and SclI show some variation in their collagen-like regions compared to their *Sz*H70 counterparts (70 to 82% amino acid sequence identity) but are most variable in the V domain (19 to 60% aa sequence identity). FneF, SclH and SclD (*Se*4047) have diverse collagen-like regions (32 to 74% amino acid sequence identity) but more conserved V regions (89 to 100% aa sequence identity) compared to their *Sz*H70 homologues. SclD (*Se*4047) differs from SZO17540 (*Sz*H70) only in the number of GXY-motifs but in all other cases the collagen-like regions vary in both primary sequence and the number of GXY-motifs. 8 out of the 10 *Se*4047 CDSs (9 Scls and FneE) have 6 to 87 fewer GXY-motifs relative to their *Sz*H70 homologues, whilst 2 (SclF and SclI) have an extra 6 to 9 GXYs (on average the *Se*4047 CDSs have 22 fewer GXY-motifs) (Table 2.5).

A number of the *Sz*H70 Scls have a much closer homologue in *Sz*MGCS10565 than in *Se*4047, which may indicate subspecies-specific adaptations in these proteins. For example, SclG and SclE (*Se*4047) are more divergent from their homologues in *Sz*H70 and *Sz*MGCS10565 (Table 2.5). In addition, SZO00830 and SclZ.1 have no homologous CDS at the same genome location in *Se*4047, and only a truncated *Se*4047 collagen binding protein FneD (which has no cell wall anchor) is present in what appears to be a highly

variable region of the *S. equi* subspecies genomes encoding polymorphic putative cell wall-anchored proteins in *Sz*H70 (SZO05350) and *Sz*MGCS10565 (SclZ.9) (Table 2.5).

Potential integrin binding sequences within the collagen-like regions of the Scls have been highlighted in Table 2.5. Variation within the amino acid residues of these collagen-like domains has led to changes in the number and type of putative integrin binding motifs. For example, SclG (*Se*4047) contains one RGD, 2 KGD and one GKPGER whilst its homologues in *Sz*H70 and *Sz*MGCS10565 have one GPP and 23 to 30 KGD motifs.

2.5 Discussion

Available evidence suggests that *S. equi* evolved from an ancestral *S. zooepidemicus* strain, a near commensal bacterium largely associated with subclinical inflammatory airway disease (IAD) in young horses (Wood *et al.*, 2005a) but also able to cause opportunistic disease in a range of hosts. Given the close genetic similarity between these 2 streptococci, comparative genomics was used to identify differences between the subspecies with a view to distinguishing key determinants of their differing pathogenic properties. This analysis has revealed the forces of gene gain, gene loss and gene change that have all contributed to the evolution of *S. equi* into a highly specialised host-restricted bacterial pathogen.

2.5.1 Pathogenic specialization and gene gain

The increased size of the *Se*4047 genome compared to the genome of *Sz*H70 is due to the acquisition of a large number of MGEs. The gain of these regions by a progenitor may have opened up new pathogenic niches, and been critical in the emergence of *S. equi.* Unlike *Sz*H70, *Se*4047 is polylysogenic, containing 4 prophage. The transductional acquisition of prophage plays an important role in the evolution of many pathogenic

bacteria (Brussow *et al.*, 2004). Cargo genes carried by prophage can increase the survival fitness or enhance niche adaptation of the lysogen (Beres & Musser, 2007; Brussow *et al.*, 2004). φSeq2 has introduced into the *Se*4047 genome a CDS that shares 98% amino acid sequence identity with the secreted phospholipase A_2, SlaA, of *S. pyogenes* M3 MGAS315 (Brussow *et al.*, 2004). SlaA is related to a potent toxin (textilotoxin) made by the Australian brown snake and has enzymatic activity against several phospholipid groups (Nagiec *et al.*, 2004). For example, SlaA cleaves and releases arachidonic acid, a powerful mediator of the inflammatory cascade. SlaA is known to contribute to streptococcal virulence and its recent acquisition by *S. pyogenes* M3 (in approximately 1987) was associated with increased morbidity and mortality (Brussow *et al.*, 2004). Deletion of *slaA* reduced the virulence of *S. pyogenes* in a mouse intraperitoneal infection model and severely compromised its ability to colonise the upper respiratory tract of a macaque model of pharyngitis (Sitkiewicz *et al.*, 2006). An impaired ability to adhere to and kill pharyngeal epithelial cells was also demonstrated by the *ΔslaA* mutant strain in comparison to the wild-type strain (Sitkiewicz *et al.*, 2006). *slaA* was present in all strains of *S. equi* and was widely distributed amongst *S. zooepidemicus* STs (31% of isolates examined) suggesting that phage lysogeny is also a feature of a subset of the *S. zooepidemicus* population (Figure 2.6). Interestingly, 7 *S. zooepidemicus* ST groups included isolates that had *slaA* as well as isolates that were negative for this gene. This observation raises the possibility of recent horizontal gain of *slaA* by some members of these ST groups. Of particular note was the observation that *S. equi* CF32, which was isolated from a horse with strangles during 1981, contained *slaA*. This isolate predates all *slaA* positive isolates of *S. pyogenes* (Beres *et al.*, 2002), and it is possible that *slaA* in the *S. pyogenes* gene pool gene may have arisen via phage-mediated horizontal transfer from a *slaA*-containing strain of *S. zooepidemicus* or *S. equi*, although the precise evolutionary origins remain unclear.

In support of the important role that these phospholipase toxins may play in colonisation and virulence of *S. equi* and *S. zooepidemicus*, a gene encoding a second putative phospholipase A_2 toxin, SlaB, sharing 70% amino acid sequence identity with SlaA of *S. pyogenes* in the genomes of *Se*4047 and *Sz*H70 was identified. *slaB*, a feature of all strains of *S. equi* and *S. zooepidemicus* (Figure 2.6), is likely to have been gained some time ago in the evolutionary history of these 2 subspecies. *slaB* was probably introduced via phage transduction as a hypothetical prophage gene (SEQ2154/SZO18660) is still present adjacent to *slaB* in both genomes. This additional gene may increase the dosage of phospholipase A_2 produced by *S. equi* and some strains of *S. zooepidemicus* and increase host cell cytotoxicity. Overcoming the epithelial cell barrier may be important during the rapid invasion of *S. equi* in the early stages of infection. Certainly, SlaA is secreted very early in the interaction of GAS with host material (Banks *et al.*, 2003; Nagiec *et al.*, 2004; Shelburne *et al.*, 2005). A continued host inflammatory response has been observed in the GPs of horses with prolonged carriage of *S. equi*. The activity of phospholipases and/or other proinflammatory mediators may contribute to this pathology and cell damage may, for example, release nutrients that contribute to the prolonged survival of *S. equi* in this site. Increased phospholipase A_2 activity may also expand the population size of *S. zooepidemicus* isolates colonising the equine respiratory tract thereby promoting their persistence and transmission.

φSeq3 carries CDSs encoding the superantigens SeeL and SeeM, which share 97% and 96% amino acid sequence identity with SpeL and SpeM of *S. pyogenes* MGAS8232, respectively (Alber *et al.*, 2005; Proft *et al.*, 2003). These mitogenic proteins belong to an increasing family of related superantigenic exotoxins produced mostly by *S. pyogenes* and *S. aureus*. By simultaneously interacting with MHC class II molecules of antigen presenting cells and with T cells at specific motifs of the T cell receptor, these exotoxins simulate T lymphocyte proliferation causing a massive systemic release of pro-

inflammatory cytokines (Fraser & Proft, 2008). Streptococcal exotoxins are implicated in severe group A streptococcal infections of man including scarlet fever and streptococcal toxic shock syndrome (Fraser & Proft, 2008). The *speL* gene has been recently acquired by M3/T3 *S. pyogenes* isolates in Japan (sometime between 1973 and 1992) that have caused more than 10% of streptococcal toxic shock-like syndrome cases reported from 1992 to 2002 (Ikebe *et al.*, 2002). In addition, the presence of the *speL* gene is highly associated with *S. pyogenes* M89, a serotype linked to outbreaks of rheumatic fever in New Zealand (Proft *et al.*, 2003). Such extreme outcomes of superantigen function are unlikely to benefit the bacterium but an advantage is almost certainly gained during the early stages of infection when the superantigen induces a localized anesthesia of the innate immune responses that are essential for a rapid clearance of the invading bacterium. A similar function may be particularly important to *S. equi* during its brisk invasion of the equine respiratory tract.

It is particularly striking that *S. equi* appears unaffected by high numbers of neutrophils in abscessing LNs, which would normally be able to rapidly kill invading bacteria. This suggests that *S. equi* may be able to misdirect the equine immune response and actually harness its power to enhance abscess formation. The genes encoding SeeL and SeeM were present in all strains of *S. equi* but only 4 of 140 isolates of *S. zooepidemicus* tested (Figure 2.6). Interestingly, these *S. zooepidemicus* isolates represented 3 unrelated STs (ST-106, ST-118 and ST-120) recovered from the same outbreak of equine respiratory disease in 1996. In addition, *seeL* and *seeM* were absent from other representatives of ST-106 and ST-118. *S. equi* CF32 also contained these superantigen genes, and predates SpeL- and SpeM-producing strains of *S. pyogenes* (Ikebe *et al.*, 2002), providing further evidence that *S. equi* and *S. zooepidemicus* act as reservoirs of virulence genes that may be transferred by lateral gene transfer events.

φSeq4 contains cargo CDSs encoding the previously described superantigens SeeH and SeeI, which share 98% and 99% amino acid sequence identity with SpeH and SpeI, respectively (Artiushin *et al.*, 2002). Interestingly, φSeq4, is very closely related to φMan3 of *S. pyogenes* Manfredo (Figure 2.7, Matt Holden) (Holden, 2009). *seeH* and *seeI* were present in all strains of *S. equi* but have not yet been detected in any strains of *S. zooepidemicus* examined. Re-circularised φSeq4 has been detected by PCR of phage particles purified from culture of *Se*4047 (Holden, 2009). These data suggest that the acquisition of φSeq4 by *S. equi*, possibly originating from a strain of *S. pyogenes*, may have been a very recent event that could have influenced the emergence of *S. equi*. Recombinant SeeI, but not SeeH, was recently shown to induce a strong dose dependent proliferative response in equine CD4+ T lymphocytes and synthesis of interferon-γ (Romain Paillot, unpublished data). In agreement with this result, deletion of *seeI*, but not *seeH*, by allelic replacement in *Se*4047, resulted in a reduction in mitogenic activity elicited by culture supernatant (Romain Paillot, unpublished data). In contrast, earlier work showed strong mitogenic responses by horse PBMCs to both SeeI and SeeH, although only the former was pyrogenic in horses (Artiushin *et al.*, 2002). Since only femtomolar quantities of SeeI elicit a mitogenic response from PBMCs, it is possible that cross-contamination confounded these earlier results. Alternatively, PBMCs from individual horses may respond differently. Although not pyrogenic in horses, SeeH was able to induce pyrexia in rabbits (Artiushin *et al.*, 2002) and SpeH, like SpeI, was mitogenic for human PBMCs (Brouillard *et al.*, 2007; Proft *et al.*, 1999). Overall, these data suggest that the recently acquired SeeH superantigen is non-functional at least in some horses and therefore not adapted to the equine host.

Phage can also improve the competitive fitness of the host strain through spontaneous induction to the lytic phase in a small number of host cells, resulting in the destruction of rival bacteria (Bossi *et al.*, 2003). The effect is exerted at the population level since the

individual bacteria that suffer induction are presumably killed in the process; a form of altruistic suicide. Phage repressor and superinfection exclusion functions in the lysogens provide immunity against such lytic infection (Bossi *et al.*, 2003). From this perspective, *S. equi* has a formidable arsenal of weapons. Both φSeq1 and φSeq4 were recently shown to revert to the lytic cycle following mitomycin C treatment and although phage particles were not detected from φSeq2 and φSeq3, these prophage appear intact and may be induced by alternative stimuli (Holden, 2009). *Salmonella enterica* serovar *Typhimurium* lysogens are able to out-compete non-lysogenic strains through the release of low titers of phage and this effect is even more dramatic when the lysogenic strains release more than one phage type (Bossi *et al.*, 2003). Release of phage from *S. equi* may result in killing of susceptible bacteria, such as *S. zooepidemicus*, which may compete to colonise the epithelium of the equine nasopharynx, thereby resulting in more efficient attachment of *S. equi* and its invasion of the lymphatic system. The persistence of *S. equi* in the GP of horses, particularly in the abscence of chondroids, may be aided by such competitive measures; *S. zooepidemicus* isolates are occasionally cultured from this site in horses.

Such a symbiotic phenomenon may have little cost to the lysogenic host strain and benefits the phage by enabling the spread of viral DNA. Opportunities for reassortment among different bacteriophage are also likely to arise and virulence cassettes could be further mobilised by this process. The diversity of prophage sequences identified in the genomes of streptococci and other pathogens is considerable and similarly, the prophage found in *Se*4047 showed only limited mosaic similarity with each other (Holden, 2009). Interestingly, the prophages of *S. equi* showed most extensive similarity with prophage from *S. pyogenes* suggesting commonality in the phage pool of these pathogens (Holden, 2009). This agrees with the likely horizontal transfer of *slaA*, *seeL*, *seeM*, *seeH* and *seeI* between the group C and group A streptococci. The 25 strains of *S. zooepidemicus* that demonstrated mitogenic activity but lacked *S. equi* superantigen genes probably contain

genes encoding other *S. pyogenes* superantigens or novel genes that represent an additional reservoir of as yet uncharacterised exotoxins. Interestingly, a significantly higher proportion of respiratory isolates of *S. zooepidemicus* had mitogenic activity (30%, n = 73) compared to other *S. zooepidemicus* isolates (9%, n = 65, P = 0.0027, Fischer's exact test), which suggests that superantigens contribute to the ability of these strains to colonise the respiratory tract.

The absence of prophage in the *Sz*H70 (and *Sz*MGCS10565) genome, and low frequency of phage associated superantigens in the screening of *S. zooepidemicus* strains, is in stark contrast to *S. equi*. One explanation for the lack of prophage in *S. zooepidemicus* is that systems exist in naturally transformable streptococci that provide resistance to uptake and incorporation of foreign DNA and may co-incidentally prevent stable prophage integration (Beres *et al.*, 2008). Prophage are not present in any of the 7 sequenced genomes of the naturally transformable streptococci, including *S. pneumoniae*, *S. gordonii*, *S. sanguinis*, or *S. mutans* (Ajdic *et al.*, 2002; Hoskins *et al.*, 2001; Lanie *et al.*, 2007; Tettelin *et al.*, 2001; Vickerman *et al.*, 2007; Xu *et al.*, 2007). Natural transformation by *S. zooepidemicus* has not yet been assessed but the presence of bacteriocin/competence gene homologues in the genome of *Sz*MGCS10565 points to the possibility that at least some strains of this subspecies might be naturally competent (Beres *et al.*, 2008). Recombination is also a predominant feature of the highly diverse *S. zooepidemicus* population (Webb *et al.*, 2008) and other bacterial species that are naturally competent (Alm *et al.*, 1999; Nassif *et al.*, 1993; Read *et al.*, 2000). *Se*4047 lacks 9 intact putative competence genes (Appendix Table A.2) that are present in *Sz*H70 and *Sz*MGCS10565, which could provide an explanation for the polylysogenic nature of *Se*4047. 2 of these putative competence genes were also absent from a strain of *S. zooepidemicus* (*Sz*1775), which is more closely related to *Se*4047 than *Sz*H70 (Andrew Waller, unpublished observations). *Se*4047 may therefore have evolved from a strain of *S. zooepidemicus* already deficient in competence formation

and functional redundancy in the remaining competence genes may have led to their subsequent loss. Integration of φSeq3 into a putative late competence protein (SEQ1725: ATP-binding DNA helicase/translocase) could have contributed to a redundancy in competence genes.

An alternative explanation of the proliferation of prophage in *S. equi* can be found in the genome comparison between *Sz*H70 and *Se*4047. In the *Sz*H70 genome a locus containing a CRISPR array and CRISPR-associated genes was identified (Figure 2.8), which has been deleted from the *Se*4047 genome due to recombination between ISSeq11 elements. CRISPR arrays are composed of direct repeats that are separated by similarly-sized non-repetitive spacers. These arrays, together with a group of associated proteins, confer resistance to phage directed by sequence similarity between the spacer regions and the phage in question, possibly via an RNA-interference-like mechanism (Barrangou *et al.*, 2007; Sorek *et al.*, 2008). In response to phage infection, bacteria integrate new spacers that are derived from phage genomic sequences, which results in CRISPR-mediated phage resistance (Barrangou *et al.*, 2007). Because new spacer units are added at the leader-proximal end of the array, a historical record of phage attack is revealed upon examination of the spacer sequences. *Sz*H70 appears to have resisted infection by prophage containing CDSs similar to those present in φSeq1 and φSeq3 of *Se*4047 as well as other prophage of *S. pyogenes* (φMan.2 and 370.3). Particularly interesting, is a region in the SeeM superantigen CDS that has been stored in the *Sz*H70 CRISPR array (spacer 15). Without such a defence mechanism in place, this superantigen could have been gained by *Sz*H70, which may have altered its pathogenic profile. A similar CRISPR system is present in *Sz*MGCS10565, however, with the exception of one spacer sequence, the number and sequence of spacers differs in this *S. zooepidemicus* strain refecting likely differences in phage exposure in the respective environments of these 2 *S. zooepidemicus* strains. Additional, distinct CRISPR arrays are also present in *Sz*MGCS10565 (Beres *et al.*, 2008)

but absent from *Sz*H70 and *Se*4047. The CRISPR loci of *Sz*H70 and *Sz*MGCS10565 may assist the development of resistance to circulating phage and maintain genome integrity. A CDS in the CRISPR region of *Sz*H70 (SZO14370) was present in 93% of *S. zooepidemicus* isolates examined by PCR, but was absent from all strains of *S. equi* tested (Figure 2.6). Deletion of the CRISPR locus from the ancestor of *Se*4047 is likely to have resulted in increased genome instability and illustrates that in some circumstances gene loss may in turn influence the subsequent rate of gene gain. As mentioned above, *Sz*1775, which is more closely related to *Se*4047 than *Sz*H70, has lost 2 putative competence associated genes; however, this strain of *S. zooepidemicus* does have a CRISPR locus and no prophage in its genome (Andrew Waller, unpublished data).

Both the *Sz*H70 and *Se*4047 genomes contain different ICE regions. This type of MGE has been shown to be widely distributed (Burrus *et al.*, 2002a), and associated with the transfer of a diverse range of functions. The 4 ICEs in the *Se*4047 and *Sz*H70 genomes all carry distinct cargo, although 3 of these elements have related conjugation modules that are similar to those encoded in an ICE that confers erythromycin resistance in *S. pyogenes* strain MGAS10750 (Beres & Musser, 2007) and in 2 conjugative transposons of *C. difficile* strain 630 (Sebaihia *et al.*, 2006). These group A and group C streptococci may therefore share a common ICE pool and may also acquire these MGEs from more distantly related bacteria. ICE MGAS10750-RD.2 was not present in other sequenced *S. pyogenes* strains (Beres & Musser, 2007) but the erythromycin resistance gene, *ermA* (or *erm*(TR)), carried by this ICE is a common determinant of macrolide resistance in *S. pyogenes* (Robinson *et al.*, 2006) and has been identified in other *Streptococcus* species such as *S. agalactiae* (Marimon *et al.*, 2005) and group G streptococci (Woo *et al.*, 2003). In addition, conjugal transfer of *erm*(TR) has been demonstrated between strains of *S. pyogenes* and other species such *Enterococcus faecalis*, *Listeria innocua* and *Peptostreptococcus magnus* (Giovanetti *et al.*, 2002; Reig *et al.*, 2001). Akin to the

exchange of toxins between distinct prophage, it seems likely that virulence determinants may also be transferred between different ICEs. *erm*(TR) was recently discovered in *S. pneumoniae* AP200, encoded on a transposon/prophage remnant chimera Tn*1806,* which has partial similarity to ICE MGAS10750-RD.2 (Camilli *et al.*, 2008).

One of the ICEs in the *Se*4047 genome, ICE*Se2*, contained CDSs (SEQ1233-SEQ1246) with similarity to the NRPS systems of *Clostridium kluyveri* and *Yersinia* sp. that produce an unnamed siderophore (Seedorf *et al.*, 2008) and the ferric iron-binding siderophore yersiniabactin (Bobrov *et al.*, 2002), respectively. Work described in Chapter 3 demonstrates that the *S. equi* NRPS operon is required for the production of an undefined secreted molecule, provisionally named equibactin, which enhances the ability of *S. equi* to acquire iron. Siderophore biosynthesis has not previously been identified in any streptococci (Eichenbaum *et al.*, 1996). However, homologues of SEQ1246 and SEQ1243 (present as a pseudogene) are in the genome of *S. agalactiae* NEM316 serotype III, suggesting that a locus with similarity to the *S. equi* NRPS operon may have been important to this organism at some time.

The ICE*Se2* locus was present in all of the *S. equi* isolates, but in none of the diverse collection of *S. zooepidemicus* isolates examined (Figure 2.6). Given the importance of iron acquisition to other streptococcal pathogens (Brown & Holden, 2002), the acquisition of ICE*Se2* may have contributed significantly to the increased pathogenesis of this *Streptococcus*. For example, the more efficient acquisition of iron could enhance the ability of *S. equi* to generate LN abscessation, which is critical to the establishment of long term carriage and vital to the success of this bacterium. It is intriguing to note that the production of yersiniabactin by *Y. pestis* is essential to its virulence (Bearden *et al.*, 1997). It will be important to determine the contribution of ICE*Se2* to the formation of abscesses in the LNs of horses.

A facet of the *Se*4047 genome suggestive of recent niche adaptation is the large increase in the number of IS elements relative to *Sz*H70 (*Sz*H70 contains 30 whereas *Se*4047 contains 73; Table 2.3, Matt Holden). In particular, there appears to have been an expansion of the IS3-family IS element, ISSeq3: the *Se*4047 genome contains 40 copies of ISSeq3 whereas *Sz*H70 contains 4 (ISSzo3). An expansion of IS elements has been observed in several host-restricted pathogens, which have recently evolved from generalist ancestors (Nierman *et al.*, 2004; Parkhill *et al.*, 2003). An evolutionary consequence of niche transit is hypothesised to be that many genes become dispensable, allowing increased inactivation. Niche change is also associated with significant evolutionary bottlenecks, which will be enhanced by repeated acquisition of MGEs. This leads to small effective population sizes, resulting in lower efficiency of selection, which in turn allows gene mutation and expansion of IS elements through accelerated genetic drift. A corollary of the IS proliferation has been the loss of genes by deletion (Nierman *et al.*, 2004): several examples of gene loss (discussed below) probably occurred through insertion and recombination between IS elements (Appendix Table A.2).

2.5.2 Functional loss

*Se*4047 has 58 partially deleted genes and 78 pseudogenes, compared with 62 and 29 respectively in *Sz*H70 (Figure 2.2B, Matt Holden and Appendix Table A.2). In particular, *Se*4047 is enriched for mutations associated with catabolic metabolism, transport, and the cell envelope. Such gene loss is typical of other host-restricted bacteria that have evolved from versatile ancestors (Nierman *et al.*, 2004; Parkhill *et al.*, 2003). The loss of ancestral functions appears to have played a seminal role in the evolution of *S. equi*, resulting in a refinement of its nutritional capabilities, and its host-cell interactions.

Genome-wide signature tagged mutagenesis and/or expression microarray analysis has linked the virulence of streptococcal pathogens to carbohydrate metabolism, including genes involved in lactose metabolism (Shelburne *et al.*, 2008a). Interestingly, carbohydrate catabolism in streptococci plays an important role in colonisation of mucosal surfaces (Shelburne *et al.*, 2008a; Shelburne *et al.*, 2008b). Comparison of the *Sz*H70 and *Se*4047 genome sequences revealed deletion of genes necessary for lactose, sorbitol and ribose fermentation in the latter. This is the first time that the molecular basis for the diagnostic differentiation of *S. equi* from *S. zooepidemicus* (Bannister *et al.*, 1985) based on carbohydrate fermentation has been revealed. An inability to ferment these 3 sugars is a consistent feature of *S. equi* strains; whilst all strains of *S. zooepidemicus* were able to utilise lactose and all but one used sorbitol during growth. 15 out of 140 *S. zooepidemicus* isolates failed to ferment ribose and 13 concomitantly lacked the *rbsD* gene (Figure 2.6). These data highlight lactose as the most reliable biochemical indicator for discrimination of the 2 subspecies. Specialization of *S. equi* has probably rendered these pathways redundant, resulting in their loss. Carbohydrate usage may enhance the colonising potential of *S. zooepidemicus* and an ability of this subspecies to inhabit broader host environments where alternative carbohydrate sources are available. For example, lactose utilisation would be a useful attribute in strains of *S. zooepidemicus* associated with mastitis and zoonotic outbreaks linked to contaminated dairy products.

In addition, *Sz*H70 (and *Sz*MGCS10565) encode a putative β-galactosidase, a sorbose superfamily and cellobiose sugar PTS, all of which are disrupted in *Se*4047 (SEQ0866, SEQ0871, SEQ1613, Appendix Table A.2). A branched chain amino acid transporter SZO14690 present in all *S. zooepidemicus* isolates but absent from *S. equi* isolates (Figure 2.6) might contribute similarly to the more versatile nature of *S. zooepidemicus*. *S. equi*, having acquired many pro-inflammatory factors, might rely more heavily on host cell breakdown in order to acquire its energy sources. When environmental glucose levels are

low, GAS increase their production of virulence factors, such as SpeB, a streptococcal pyrogenic exotoxin with cysteine protease activity that could contribute to host damage and release nutrients in that way (Shelburne *et al.*, 2008b).

HA lyases are secreted enzymes that degrade hyaluronic acid and chondroitins potentially facilitating invasion by bacteria and their toxins (Hynes & Walton, 2000). A HA lyase, present in all isolates of *S. zooepidemicus* examined, has been inactivated in all *S. equi* strains and *S. zooepidemicus* ST-57 (Figure 2.6). *Se*4047 has acquired a different HA lyase encoded on a prophage, which, like other examples of this type of phage-encoded enzyme is likely to have much lower activity and a reduced substrate range (Baker *et al.*, 2002; Lindsay *et al.*, 2009) than orthologues of the *Sz*H70 HA lyase (SZO06680) (Pritchard *et al.*, 1994) and may provide an explanation for why *S. equi* infection rarely progresses beyond the lymphatic system.

Reduced HA lyase activity provides a possible explanation as to why *S. equi* maintains high levels of HA capsule and in agreement with this, the ST-57 isolate of *S. zooepidemicus* that tested PCR negative also maintained high levels of capsule. HA lyase production contributes to capsule breakdown during periods of reduced capsule synthesis in GAS (Starr & Engleberg, 2006). The mucoid colony phenotype of *Se*4047, *Sz*H70 and the ST-57 isolate was shown to be due to hyper-encapsulation as these all had a non-mucoid phenotype when grown on plates containing hyaluronidase (Holden, 2009). Increased levels of capsule enhances resistance to phagocytosis (Anzai *et al.*, 1999b; Wessels *et al.*, 1991; Wibawan *et al.*, 1999) (unpublished data), an important feature of *S. equi*, which is demonstrated by the accumulation of long chains of extracellular streptococci surrounded by large numbers of degenerating neutrophils in the LN abscesses. More capsule could also reduce adhesion to the mucosal surface (J Slater, unpublished data) (Bartelt & Duncan, 1978; Srivastava & Barnum, 1983). As HA is a major constituent

of the ECM, *S. zooepidemicus* may use HA lyase to aid nutrient release (Starr & Engleberg, 2006) and even use its own capsule as an energy source during periods of starvation.

The ability of *S. equi* to survive despite considerable neutrophil recruitment may have lessened the need of this bacterium to maintain an active C5a peptidase. The C5a peptidase enzymes of group A (ScpA) and B (ScpB) streptococci specifically cleave and inactivate the complement system chemotaxin C5a, thereby reducing the trafficking of neutrophils to the site of infection (Chmouryguina *et al.*, 1996; Cleary *et al.*, 1992; Hill *et al.*, 1988; Ji *et al.*, 1996; Wexler *et al.*, 1985). The C5a peptidase homologue encoded by *Se*4047 (SEQ1077), which shares 35 to 37% amino acid sequence identity with ScpA/ScpB, has a frameshift mutation after codon 58 and an early stop at codon 94 that truncates this protein (Table 2.5, Appendix Table A.2D). Orthologues in *Sz*H70 (SZO10150) and *Sz*MGCS10565 (Sez_0946) remain intact. A novel factor H binding protein Se18.9 (SEQ0235) (Tiwari *et al.*, 2007) that is produced by all *S. equi* strains (Figure 2.6) may have allowed some redundancy in bacterial mechanisms aimed at subverting the complement system. In addition to this, another chemokine protease (SEQ0563), which shares 61% shared amino acid sequence identity with the GAS interleukin-8 ScpC (Edwards *et al.*, 2005; Zinkernagel *et al.*, 2008) is present and intact in *Se*4047. It is also possible that SEQ1077 gene decay has resulted from positive selection and a drive to enhance abscess formation, augmented by the production of superantigens.

Ess specialised secretion systems have been identified in *M. tuberculosis* and *S. aureus* and shown to trigger cell-mediated immune responses, including interferon-γ production, that play an important role in virulence (Burts *et al.*, 2005). The *Sz*H70 genome contains a cluster of 8 genes with similarity to the *ess* of *Sz*MGCS10565 (Beres *et al.*, 2008) and *S. aureus* (Figure 2.9) (Burts *et al.*, 2005; Burts *et al.*, 2008; Sundaramoorthy *et al.*, 2008).

The *Se*4047 genome lacks *esxA*, *esaA* and part of *essA* most likely as a result of recombination between ISSeq3 elements, one of which is still present at this site (SEQ0575-SEQ0574). *esaB* is also a pseudogene in *Se*4047. PCR analysis showed that none of the 26 isolates of *S. equi*, but 138 of 140 *S. zooepidemicus* isolates examined contained the *esaA* gene (Figure 2.6). This type VII secretion system is unlikely to be functional in *S. equi* since EssA and EsaB are required for the production of the secreted factors EsxA, EsxB (Burts *et al.*, 2005) and EsaC (Burts *et al.*, 2008) respectively in *S. aureus*. EssA is a predicted membrane protein that is presumed to form part of the secretion apparatus, whilst EsaB acts as a post-transcriptional negative regulator of EsaC production. Given that components of the *S. aureus* Ess promote persistent abscesses during animal infection, the finding that this system is partially deleted in *S. equi* is surprising. However, a type VII secretion system was not required for *L. monocytogenes* virulence in mice (Way & Wilson, 2005) and similar systems are also present in non-pathogenic Gram-positive and mycobacterial species (DiGiuseppe Champion & Cox, 2007). These systems may therefore have functions that are shared by pathogenic and non-pathogenic bacteria such as cell-to-cell communication (DiGiuseppe Champion & Cox, 2007). The EsaC protein is more divergent among Gram-positive species and may confer different functions in other systems. Differences in the pathogenic strategies employed by *S. equi* and *S. zooepidemicus* may also reflect the differing contributions of these alternative secretion systems to virulence. For example, *S. equi* may rely more heavily on superantigens to disrupt the host immune response.

Gram-positive bacteria classically display an array of cell wall-anchored proteins on their surface, which are attached covalently through a process mediated by sortase enzymes (Marraffini *et al.*, 2006). In many cases, these cell wall-anchored proteins have been shown to play a role in modulating host-cell interactions. The genome of *Se*4047 has

undergone noteable gene decay in its collection of surface proteins in comparison to *Sz*H70 and *Sz*MGCS10565 (Table 2.5).

S. equi strains typically bind significantly lower quantities of fibronectin than those of *S. zooepidemicus* (Lindmark *et al.*, 2001). One possible explanation for this is a one-base deletion within SEQ0375 (*fne*) that was conserved in all strains of *S. equi* examined (Lindmark *et al.*, 2001). The base deletion in *fne* results in the loss of an LPXTG surface anchor and leads to the production of a secreted product, Fne, which binds both fibronectin and collagen (Lindmark *et al.*, 2001). Interactions with fibronectin have been shown to attenuate the virulence of *S. pyogenes* in mice (Nyberg *et al.*, 2004). Similarly, reduction in the fibronectin-binding properties of *S. aureus* increases virulence in a rat pneumonia model (McElroy *et al.*, 2002) and truncation of *fne* has been proposed to increase the virulence of *S. equi* (Lindmark *et al.*, 2001). This study identifies other examples of mutation and gene loss that are likely to contribute to decreased fibronectin binding in *S. equi*. The surface protein Shr of *S. pyogenes* binds heme and transfers it to the streptococcal heme-binding protein Shp for import by the HtsABC heme transporter (Zhu *et al.*, 2008). Shr also binds fibronectin and contributes to attachment of *S. pyogenes* to epithelial cells (Fisher *et al.*, 2008). SEQ0443 encoding Shr in *S. equi* contains a frameshift mutation after codon 442 that truncates this protein. ScpB, the streptococcal C5a peptidase of *S. agalactiae* was recently also shown to bind fibronectin (Tamura *et al.*, 2006). SEQ1077 (*scpB*) is a pseudogene in *Se*4047 and encodes a truncated 93 residue protein. CDSs for another 4 cell-wall anchored proteins that potentially function as adhesins by binding host receptors or components of the ECM are also inactivated in *Se*4047. These include a SeM-like protein (SEQ0232), a collagen-like protein (SEQ0556), a Listeria-Bacteroides repeat containing protein (SEQ1307a) and a SzPSe-like protein (SEQ0566). Mostly, these are unlikely to produce functional proteins due to the introduction of a very early stop codon (SEQ0232), the lack of a signal sequence

(SEQ0566 and SEQ1307a) and/or near complete degradation (SEQ1307a). However, SEQ0566 is likely to secrete a 182 amino acid truncated product that may still retain ligand binding activity. In addition, FneD, which has collagen-binding activity, but also lacks C-terminal residues, is likely to be secreted into the extracellular milieu (Lannergard, 2006). *S. equi* has therefore evolved to secrete a number of potential and proven host ligand binding proteins through a process of gene decay. Such changes are likely to have reduced the surface bound adhesive capacity of *S. equi* and may also reduce the number of host targets available for binding, either by *S. equi* or by competing bacteria. Sfs, a secreted fibronectin binding protein of *S. equi*, absent from 15% of *S. zooepidemicus* strains, may act in a similar fashion (Lindmark & Guss, 1999). These secreted proteins may confer other functions as demonstrated by Fne, which is able to re-shape host cells through an interaction with host integrins and ECM components. This interaction has been proprosed to offer a novel mechanism to protect *S. equi* from the innate immune response (Liden *et al.*, 2008) but may also contribute to *S. equi* virulence in yet unknown ways.

One of the *Sz*H70 cell-wall anchored proteins deleted in *Se*4047 contains Listeria-Bacteroides repeat domains that are common to families of internalins of *Listeria* species (Figure 2.10). InlA, the internalin of *L. monocytogenes* (Orsi *et al.*, 2007) interacts with E-cadherin to promote invasion of *L. monocytogenes* into particular host cells (Ireton, 2007). The production of this surface protein is likely to be switched on and off through the action of a putative site-specific recombinase (SZO08550) in a manner akin to phase variation in *E. coli* or *B. fragilis* (Figure 2.11) (Abraham *et al.*, 1985; van der Woude & Baumler, 2004). This is the first potential example of recombinase regulation of surface protein production in streptococci. Such phase variation would result in a heterogeneous *S. zooepidemicus* population, in which some bacteria have a greater potential to maybe adhere to or invade certain host cells. Phase variation may also help the bacteria to evade the host immune system. Whilst this regulated surface protein was absent from all isolates

of *S. equi* (n = 26), its presence was detected in a significantly higher proportion (72%, n = 140) of *S. zooepidemicus* isolates (P < 0.0001, Fischer's exact test). SZO08550 was more prevalent in *S. zooepidemicus* isolates collected from the respiratory tract (including LN abscesses) or cases of abortion/uterine infection (77%, n = 118) compared to other *S. zooepidemicus* isolates (66%, n = 20, P = 0.002, Fischer's exact test). SZO08560 was also more prevalent in lower respiratory tract *S. zooepidemicus* isolates (including those isolated from cases of pneumonia, 100%, n = 14) than in upper respiratory tract isolates of *S. zooepidemicus* (including those associated with LN abscessation) (71%, n = 59) (P = 0.03, Fischer's exact test) and all other *S. zooepidemicus* isolates (69%, n = 124, P = 0.011, Fischer's exact test). These data suggest that the expression of SZO08560 may contribute an adaptation to particular sites of infection and in some cases increased pathogenicity. *Sz*MGCS10565 has the same locus but contains an IS element between the inverted repeats bordering the Sez_1114 (surface protein) promotor and the recombinase (Sez_1116); the consequences of this on transcription of Sez_1114 are not yet known. It was also noted that *S. zooepidemicus* isolates lacking this surface protein had a significantly higher prevalence of the pilus locus (82%, n = 39) that had been deleted from *Se*4047 (*srtC2/srtC3*, discussed below) than *S. zooepidemicus* isolates with SZO08550 (49%, n = 101, P = 0.0003, Fischer's exact test) and only rarely did *S. zooepidemicus* strains lack both (5%, n = 140, Figure 2.6). Interestingly, 5 of the 8 *S. zooepidemicus* isolates that did not contain either of these loci grouped with *S. equi* by MLST (ST-6, ST-133, ST-137 and ST-149). The remaining isolates (ST-112 and ST-122) group together by MLST, but were isolated from the equine respiratory tract and may possess an alternative mechanism(s) for host interaction.

The genome comparison has revealed clear differences between *Se*4047 and *Sz*H70 in their potential for pilus assembly. Mutation of the putative TetR-like regulator of *Se*4047 may lead to de-regulation and over-expression of CDSs from the associated putative pilus

cluster in *S. equi*. Expression of *rrgB*, which encodes the shaft subunit of the RlrA pilus in *S. pneumoniae*, is necessary in order to detect the associated accessory adhesin RrgA in the presence of capsule (Falker *et al.*, 2008). Over-expression of major pilin subunits can also result in the formation of extremely long pili (Lauer *et al.*, 2005; Scott & Zahner, 2006). Inactivation of the TetR-like regulator may therefore lead to constitutive pilus production, longer pili that could more effectively protrude through the larger capsule of *S. equi* and increase collagen binding by the adhesin Cne (Lannergard *et al.*, 2003). This putative pilus in *Se*4047 may be important during the initial stages of infection enabling the piliated bacterium to rapidly secure (Manetti *et al.*, 2007) onto pharyngeal/tonsillar epithelial cells through distant contacts. Minor pilins have been repeatedly shown to play important roles in adherence of streptococci and other Gram-positive bacteria to host tissues (Abbot *et al.*, 2007; Dramsi *et al.*, 2006; Krishnan *et al.*, 2007; Maisey *et al.*, 2007; Mandlik *et al.*, 2007; Nelson *et al.*, 2007). Interestingly, a pilus backbone of GBS contributes to the paracellular translocation of these bacteria through epithelial cells with pili observed in the intercellular space ahead of the translocating bacteria (Pezzicoli *et al.*, 2008). The route taken by *S. equi* as it moves so rapidly across the epithelial barrier is unknown, however an upregulated pilus cluster offers an exciting proposition for a process similar to the one described in GBS. Moreover, the presence of pili also stimulates the host inflammatory response (Barocchi *et al.*, 2006), which might damage the mucosal barrier and facilitate invasion of *S. equi*.

IS-mediated gene loss has removed the 2nd pilus from *Se*4047, an event that may have contributed to a reduced colonisation potential and/or the restricted host and tissue tropism of *S. equi*. The associated *srtC2/srtC3* genes were also missing from all other strains of *S. equi* tested (Figure 2.6). *S. pyogenes* pili promote bacterial aggregation, adhesion and biofilm formation (Manetti *et al.*, 2007), features that may prove important for the persistence of *S. zooepidemicus* at the mucosal surface. Pili of *S. pyogenes* also display

adhesive specificity for clinically relevant human tonsil and skin cells but not Hep-2 or A549 cells (Abbot *et al.*, 2007). In addition, tissue tropism of *C. diptheriae* for pharyngeal cells is governed by the specific minor pilins associated with one of 3 distinct pilus structures (Mandlik *et al.*, 2007).

Pilus loci in streptococci are often clustered at the same genetic locus (flanked by highly conserved genes), have a lateral mode of transmission and common ancestry but display considerable variability with some evidence of recombination facilitating the generation of variant proteins with different adhesive specificities (Bessen & Kalia, 2002; Telford *et al.*, 2006). Host immune selection may also drive change as pili are among the most immunogenic and broadly recognized GAS antigens (Manetti *et al.*, 2007). *S. zooepidemicus* shows diversity in its complement of pili, highlighted by the presence of 2 distinct pilus clusters in the same genome location in *Sz*MGCS10565 compared to *Sz*H70. Other unique clusters may account for some (or all) of the 42% of diverse *S. zooepidemicus* strains that were negative by PCR for *srtC.3* and/or *srtC.4* (Figure 2.6). Such diversification of pilus loci could play an important role in the ability of *S. zooepidemicus* strains to infect different hosts and tissues.

2.5.3 Adaptation through gene change

Adaptation to a new niche is a continuous process. Newly acquired virulence traits, along with pre-existing core genes, continue to undergo selection. SeM, the major antiphagocytic surface protein of *S. equi* was previously shown to have a unique N-terminal domain relative to its allelic variant SzM in *Sz*H70 (Kelly *et al.*, 2006). As this N-terminal region is necessary for the binding of SeM to fibrinogen and IgG, the replacement or modification of this region was proposed to represent an important evolutionary adaptation by *S. equi* that may have contributed to its increased virulence over *S.*

zooepidemicus (Kelly *et al.*, 2006). In fact similar diversification is apparent in many of the shared surface proteins of these two subspecies (Table 2.5) as well as in other proteins throughout the genomes (Appendix Table A.2C). The high level of polymorphism in surface proteins correlates with their likely role in host-pathogen interactions (host/tissue adaptation and immune evasion) and includes fibrinogen, collagen and fibronectin binding proteins as well as many others with as yet unknown function.

Collagen-like surface proteins make up a significant proportion of these diversifying cell-wall anchored proteins. A large family of Scls is a peculiar feature of *S. zooepidemicus* and *S. equi* that has presumably contributed significantly to their ability to attach, colonise and possibly invade host cells. 9 Scls (SclC-SclI, FneC, FneF) were previously identified in *S. equi* (Karlstrom *et al.*, 2004; Karlstrom *et al.*, 2006; Lannergard, 2006) and 12 putative Scls (including one unconfirmed pseudogene) were identified in the genome sequence of *Sz*MGCS10565 (Beres *et al.*, 2008). This study has identified 11 Scl CDSs in *Sz*H70. Collagen-like proteins are quite rare in prokaryotic genomes (Rasmussen *et al.*, 2003) and few bacterial genomes have multiple genes encoding collagen-like proteins. 2 collagen-like surface proteins have been identified in *S. pyogenes* (Scl1 or SclA and Scl2 or SclB), one in *S. pneumoniae* (PclA) (Paterson *et al.*, 2008) and 2 in *Bacillus anthracis* (BclA and BclB) (Steichen *et al.*, 2003; Sylvestre *et al.*, 2002; Thompson & Stewart, 2008; Waller *et al.*, 2005). Scl1 and PclA both contribute to bacterial binding and internalization into host cells (Caswell *et al.*, 2007; Lukomski *et al.*, 2000; Paterson *et al.*, 2008). Genetically diverse GAS strains display a highly polymorphic Scl1 protein (Lukomski *et al.*, 2000). 37 distinct Scl1 variants were identified in 21 M types, with most variation occurring in the N-terminal V regions and in the number and type of GXY repeats in the central collagen-like regions (Lukomski *et al.*, 2000). Comparison of Scl homologues in *Se*4047, *Sz*H70 and *Sz*MGCS10565 revealed similar genetic variation in a number of Scl

CDSs suggesting the importance of both these V and collagen-like domains in host interaction (Table 2.5).

N-terminal V regions are predicted to be involved in ligand binding. Recent studies have located low density lipoprotein- and factor H-binding activities in the V region of different GAS serotype Scl1 proteins (Caswell *et al.*, 2008b; Han *et al.*, 2006a). 3 of the *Se*4047 Scls FneC, FneE and FneF bind collagen and/or fibronectin (Lannergard, 2006). Variants of Scl1 and Scl2 demonstrate clear differences in their ligand-binding potential (Pahlman *et al.*, 2007) and the same may apply to those Scl types specific to *Se*4047, *Sz*H70 or *Sz*MGCS10565.

By mimicking the structure of mammalian collagen, collagen-like regions of Scls may also be recognised by eukaryotic integrins that normally bind to collagen. Sequences within vertebrate collagens that are targetted by collagen receptors include GxOGER- and GPO-motifs, whilst cryptic RGD and KGD motifs unveiled following denaturation of the triple helix (maybe via proteolytic degradation) are recognised by fibronectin/vitronectin receptors (Heino, 2007). Prokaryotes are unable to produce hydroxyproline (O) and so equivalent motifs in Scls might utilise proline (P) instead of O. Indeed, the GLPGER sequence in the collagenous domain of Scl1 acts as a functional integrin binding motif and mediates direct binding to the $\alpha_2\beta_1$ and $\alpha_{11}\beta_1$ integrins (Caswell *et al.*, 2008a). Such interactions trigger intracellular phosphorylation signalling cascades like those induced by integrin-binding of ECM components (Caswell *et al.*, 2007; Humtsoe *et al.*, 2005). Interestingly, Scl1 mediates GAS internalization by human pharyngeal epithelial cells upon binding to the $\alpha_2\beta_1$ integrin (Caswell *et al.*, 2007). All of the Scls in *Se*4047 and *Sz*H70 have multiple putative integrin-binding motifs so it is possible that interactions with host integrins via these motifs could provide a mechanism used by *S. zooepidemicus* and *S. equi* to abuse the host cell cytoskeleton dynamics and facilitate colonisation and/or

invasion. Multiple cell types appear to recognize and bind Scl1 via their collagen receptors, although their relative responses may be dependent on the dominance of certain integrins and their binding affinity for the GLPGER motif (Caswell *et al.*, 2008a). The differences observed in the number, type and order of potential integrin binding motifs in *Se*4047, *Sz*H70 and *Sz*MGCS10565 Scl homologues (Table 2.5) could therefore influence the different tissue types targeted by the *S. equi* and *S. zooepidemicus* subspecies, as well as their colonisation and invasion potential.

Diversification in the collagenous region of Scls may serve another purpose. Structural studies demonstrated that several Scl1 and Scl2 variants share a common 'lollipop-like' domain organization with stalks made of the collagen-like region and the globular head formed by the V region (Han *et al.*, 2006b; Xu *et al.*, 2002). The triple-helix structure may help to present the presumed ligand binding region (V domain) away from the bacterial surface. Differences in the number of GXY-motifs between homologues of Scls in *Se*4047, *Sz*H70 and *Sz*MGCS10565 may therefore contribute to this stalk function and the accessibility of protein domains to the host cells. Interestingly, the adhesive contribution of PclA in *S. pneumoniae* was only apparent in the presence of a capsule and may be related to its large size and protrusion beyond the capsule (Paterson *et al.*, 2008). Contrary to this, *Se*4047 is hyper-encapsulated *in vitro* but mostly has Scl proteins with shorter collagen-like regions in comparison to *Sz*H70 (and *Sz*MGCS10565). Interestingly, differences in the length of proline-rich C-terminal regions (and the number of proline-rich repeats) demark many of the *Se*4047 and *Sz*H70 cell-wall anchored proteins, particularly those without collagen-like domains (Table 2.5). These changes are also likely to alter the presentation of the ligand-binding site away from the bacterial cell surface.

Since *Sz*H70 (ST-1) and SzMGCS10565 (ST-72) are only distantly related to *S. equi* in comparison to other *S. zooepidemicus* subtypes, their differences in amino acid sequence of

cell-wall anchored proteins may correspond to variation that exists amongst the *S. zooepidemicus* population rather than differences that delineate the 2 subspecies. For example, 3 Fnz variants have now been identified in 4 *S. zooepidemicus* strains (*Sz*H70, *Sz*MGCS10565/VTU211 and Z5) (Hong, 2005; Lindmark *et al.*, 1996; Lindmark *et al.*, 2001) and one of these encoded by *S. zooepidemicus* strain Z5 is nearly identical to Fne in *S. equi*. These Fnz alleles are well conserved in their C-terminal fibronectin-binding regions but differ significantly in their N-terminal regions (34 to 47% amino acid sequence identity). Fibronectin-binding activity has been demonstrated in the N-terminal regions of Fnz (VTU211) and Fnz (Z5) that share 47% amino acid sequence identity (Hong, 2005; Lindmark *et al.*, 1996; Lindmark *et al.*, 2001); the latter was also shown to bind collagen. Further analysis of surface proteins in strains of *S. zooepidemicus* that are more closely related to *S. equi* may shed more light on changes that contribute to the increased pathogenicity of *S. equi*.

Chromosomal rearrangements have also contributed to gene change between the *Se*4047 and *Sz*H70 genomes. 2 inversions around the origin of replication differentiate the global gene order of *Se*4047 from *Sz*H70. Such reciprocal inversions around the origin and terminus are the most common form of chromosomal rearrangement between related bacteria (Eisen *et al.*, 2000; Tillier & Collins, 2000) and have been hypothesised to be caused by either direct recombination between replication forks (Tillier & Collins, 2000) or the fact that only those regions of DNA close to the replication forks are unpackaged and available for recombination (Deng *et al.*, 2002). Neither inversion appears to have separated any flanking genes or operonic structures that had not been previously interrupted by insertion events, although a 2nd *hasC* paralogue has been translocated to the opposite replichore in *Sz*H70 so that it is no longer 10 CDSs downstream of the gene cluster associated with capsule synthesis. *hasC* encodes uridyltransferase, an enzyme required for HA capsule production (Blank *et al.*, 2008).

Intra-replichore inversions are less common possibly because genes are transcribed in the opposite direction to the movement of the replication fork and collisions may occur between the replication and transcription machineries (Fraser, 2004). Also, genes moved closer to the origin of replication may be present in more copies per cell than genes moved closer to the terminus, as chromomal replication occurs in advance of cell division. Such a gene dosage effect will likely influence specific expression levels in the cell (Fraser, 2004). However, a 14 kb intra-replichore inversion has occurred in *Se*4047 between the 2 *hasC* paralogues that has rearranged the genes associated with capsule production (Blank *et al.*, 2008) (Figure 2.13) and may explain why *S. equi* produces such high levels of HA capsule. This intra-operon inversion has replaced the original *hasC* operon gene (SEQ0289) with the duplicated paralogue (SEQ0271), which has a novel 12 amino acid C-terminus and has also relocated the genes *gluM* and *pgi* into a separate operon (Blank *et al.*, 2008). *gluM* and *pgi* encode phosphoglucoisomerase and uridyltransferase respectively, which are involved in the pathway of UDP N-acetyl glucosamine synthesis, a precursor of HA (and peptidoglycan) (Blank *et al.*, 2008).

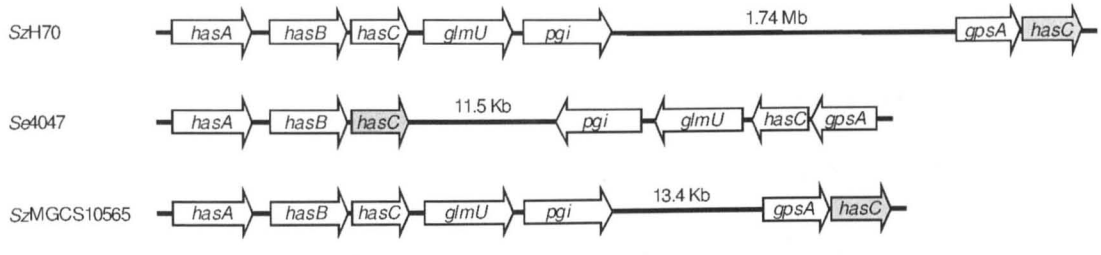


Figure 2.13 Organisation of the *has* operon

2.6 Conclusions

The comparison of the genomes of *Se*4047 and *Sz*H70 provides strong evidence that *S. equi* has passed through a genetic bottleneck during its evolution from an ancestral *S. zooepidemicus* strain. Many examples of gene loss have been identified that serve to reduce the ancestral capabilities of *S. equi* and increase the opportunity for genetic change. Different processes of gene change have accompanied the adaptation of *S. equi* to its new niche and the acquisition of new MGEs has been critical to its evolution. However, surveillance of the *S. zooepidemicus* population has identified examples of strains that did not cause strangles, but contain genes encoding phospholipase A_2 toxins and superantigens. Therefore, the key speciation event in the evolution of *S. equi* is proposed to have been the acquisition of ICE*Se2*, containing a novel NRPS involved in the acquisition of iron, which is the first of its kind to be identified in streptococci.

This study provides strong evidence for genetic exchange between *S. equi*, *S. zooepidemicus* and *S. pyogenes*, which continues to influence the pathogenicity of these important bacteria. The genetic diversity of the *S. zooepidemicus* population as measured by MLST (Webb *et al.*, 2008) suggests that further investigation of this species will be likely to identify many more genes of importance to both veterinary and human disease.

Finally, the availability of the completed genomes of *Se*4047 and *Sz*H70 has revolutionized our ability to develop new diagnostic tests and vaccines. A new strangles blood test, developed following comparison of these genomes and based on the *S. equi* specific surface antigen SEQ2190 (Figure 2.6), is available commercially and has already enhanced our ability to identify persistently infected horses. The *S. equi*-specific *eqbE* gene of ICE*Se2* has formed the basis of an improved commercial quantitative PCR test for strangles (Chapter 5). Biochemical diagnostic assays used at the AHT for the discrimination of *S. equi* and *S. zooepidemicus* have been modified to include lactose

fermentation in order to reduce the occurrence of false positive *S. equi* results. The identification of new *S. zooepidemicus*-specific DNA targets (*lacE*, SZO14690) has also improved our potential to differentiate *S. zooepidemicus* from *S. equi* using PCR. Improved knowledge of potential *S. equi* virulence factors and access to reliable infection models in the natural host, coupled with the low genetic diversity of *S. equi* revealed by MLST, should enable significant progress to be made towards the development of safe and effective vaccines and the eradication of strangles.

Chapter 3 A novel streptococcal integrative conjugative element involved in iron acquisition

3.1 Introduction

3.1.1 Integrative and conjugative elements

Horizontal gene transfer has facilitated greatly microbial evolution and rapid adaptation to new niches. The term ICE was adopted to describe all self-transmissible MGEs with predicted or demonstrated integrative and conjugative properties (Burrus *et al.*, 2002a; Burrus *et al.*, 2002b). ICEs excise from the donor chromosome to form a non-replicative circular molecule, then promote self-transfer to a new host by conjugation and subsequently insert into a new target site. This transfer can occur between genetically unrelated bacteria.

Tn*916*, identified in *Enterococcus faecalis*, was the first ICE to be completely sequenced (Flannagan *et al.*, 1994). Recently, the number of sequenced ICEs has risen considerably, particularly with the advent of genome sequencing (Burrus *et al.*, 2002a; Burrus *et al.*, 2002b; Sebaihia *et al.*, 2006). Comparisons between ICEs of low G + C Gram-positive bacteria have revealed that these elements are widespread and possess a modular organisation (Burrus *et al.*, 2002a; Garnier *et al.*, 2000; Roberts *et al.*, 2001). Each of these modules includes all of the CDSs required for a specific biological function, namely site-specific recombination, conjugation and regulation. This is not dissimilar to the combination of functional modules seen in other bacterial mobile elements such as phages and plasmids (Toussaint & Merlin, 2002) and allows these structures, even when unrelated or distantly related, to evolve by exchange of modules (Burrus *et al.*, 2002a). Consequently, a diverse set of mosaic ICEs are emerging; most encode a tyrosine recombinase (like the majority of integrated prophage) but a small number have been

identified that utilise a serine recombinase to mediate site-specific recombination (Burrus *et al.*, 2002b; Wang *et al.*, 2000).

ICEs have played an important role in the distribution of antibiotic resistance loci. Other functions encoded by ICEs include toxic aromatic compound degradation pathways, sucrose utilisation, bacteriocin biosynthesis, heavy metal resistance, type-III and type-IV secretion systems, capsule antigen and a metalloprotease virulence factor (Burrus & Waldor, 2004; Franco, 2004). A novel ICE in the ECOR31 strain of *E. coli* carries the high pathogenicity island (HPI) encoding the yersiniabactin siderophore system, which is critical to the virulence of *Yersinia sp.* and is widely distributed among *E. coli* strains and other *Enterobacteriaceae* that cause extraintestinal infections (Schubert *et al.*, 2004).

3.1.2 Non-ribosomal peptide biosynthesis

Bacteria and fungi use large multifunctional enzymes, the so-called NRPSs, to produce peptides of broad structural and biological activity. Non-ribosomal peptides include siderophores (yersiniabactin, pyochelin, enterobactin), potent antibiotics (penicillin, vancomycin), surfactants (surfactin), antitumor (bleomycin), immunosuppressive (cyclosporin) or cytostatic agents (epothilone) (Du *et al.*, 2000; Hofle, 2009; Kleinkauf & von Dohren, 1990; Marahiel *et al.*, 1997; von Dohren *et al.*, 1997; Weber *et al.*, 1994). The biosynthetic process occurs without the usual ribosomal and post-translational machinery. NRPSs are modular assembly lines that direct product formation on a protein template; reaction intermediates are covalently bounds as thioesters on phosphopantetheine prosthetic groups (Challis & Naismith, 2004; Konz & Marahiel, 1999). Genes coding for NRPSs are organized in operons or clusters and NRPSs are composed of modules, with each module responsible for the incorporation of a specific monomer. The number and order of modules corresponds to the number and order of specific incorporated monomers

and is often termed the colinearity rule. There are however several exceptions to the colinearity rule, particularly for NRPSs that assemble siderophores (including yersiniabactin) (Gehring *et al.*, 1998c; Miller *et al.*, 2002; Yu *et al.*, 1998). Modules are further subdivided into domains, each domain catalysing a specific reaction in the incorporation of a monomer. The adenylation (A) domain selects and activates the monomer transforming it into adenylate form. A thiolation or peptidyl carrier protein (PCP) domain covalently binds the activated monomer to the synthetase. The condensation (C) domain catalyses the peptide bond formation between the residues linked onto 2 adjacent modules. Finally, the thioesterase (TE) domain, present in the final module, releases the peptide from the synthetase. The product can either be released as a linear compound or be transformed into a cyclic peptide through an intramolecular reaction. In NRPSs of iterative type, the TE domain can allow the enzyme to iterate the collinear biosynthesis several times to produce a larger than anticipated product (e.g. enterobactin biosynthesis by *E. coli* (Shaw-Reid *et al.*, 1999)). Heterocyclisation (Cy) domains substitute for C-domains in several NRPSs, particularly those involved in siderophore biosynthesis (e.g. yersiniabactin) (Crosa & Walsh, 2002). The Cy-domain has a gain of function in that, after condensation, it can cyclise cysteine, serine or threonine residues to make the corresponding thiazoline or oxazaline ring (Duerfahrt *et al.*, 2004; Konz & Marahiel, 1999; Lautru & Challis, 2004). In addition, secondary domains that allow residue modications are present in many NRPSs. Examples include epimerization, methylation, oxidation and reduction domains (Konz & Marahiel, 1999).

NRPSs use amino acids (including nonproteogenic amino acids) or hydroxyl acids as building blocks, catalysing the formation of amide or ester bonds (Donadio *et al.*, 2007). The resolution of 2 A-domain crystal structures: the phenylalanine-activating A-domain (A_{Phe}) of the gramicidin S-synthetase A (GrsA, *Bacillus brevis*) (Conti *et al.*, 1997) and the 2, 3-dihydroxybenzoate (DHB)-activating A-domain DhbE of *Bacillus subtilis* (May *et al.*,

2002) allowed the identification of critical amino acid residues involved in the coordination of the substrate activated. This led to the introduction of the non-ribosomal code, which allows the prediction of A-domain selectivity on the basis of its primary sequence (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999).

A recent study analysed sequenced bacterial genomes for the presence of NRPS systems (and polyketide synthases (PKSs)) and found these to be present in 50% of 223 genomes but not uniformly distributed within the bacterial world (Donadio *et al.*, 2007). Interestingly, these biosynthetic enzymes were rare in bacteria with genomes ≤ 3 Mb. For bacteria with genomes > 5 Mb, there was a correlation between genome size and the number of NRPS/PKS clusters, suggesting that these are luxury items added after a basic self-sustaining metabolic capacity has been met (Donadio *et al.*, 2007; Sattely *et al.*, 2008). The genera *Streptomyces*, *Bacillus* and *Pseudomonas* were among those that contained a high percentage of NRPS/PKS genes. Streptococci, on the other hand, display just 2 NRPS/PKS clusters. One in *S. mutans*, the main cause of tooth decay, encodes a hybrid PKS-NRPS system consisting of 7 modules in total, 5 of which are NRPS modules (Donadio *et al.*, 2007). The identity of the corresponding metabolite has not been reported although expression from the locus is downregulated when the organism is grown in mixed species cultures (Huang, 2008). The only other *Streptococcus* strain displaying an NRPS is *S. thermophilus* (a lactic acid bacterium often used as a starter culture in yogurt and cheese), which encodes a single monomodular NRPS (Donadio *et al.*, 2007). The study also found that relatively similar NRPS clusters were found in unrelated bacteria and in most instances, these clusters directed the synthesis of siderophores, compounds that are clearly important for growth in iron-limiting environments (Donadio *et al.*, 2007). Thus horizontal transfer has likely played a strong role in spreading these biosynthetic genes.

3.1.3 Iron acquisition

Most organisms require iron as an essential element in a variety of metabolic and informational cellular pathways. More than 100 enzymes acting in primary and secondary metabolism possess iron-containing cofactors such as iron-sulfur clusters or heme groups (Miethke & Marahiel, 2007). The acquisition of iron is therefore an essential process for all pathogenic bacteria. In mammalian hosts tissue iron is sequestered by ferritin (intracellular iron storage), transferrin (plasma), lactoferrin (various bodily fluids) or in red blood cells (Wooldridge & Williams, 1993). Bacterial pathogens access this limited iron supply through the production of cell surface receptors for ferritin, transferrin and lactoferrin, utilising heme-containing compounds, iron transporters, or by the synthesis and secretion of iron-sequestering siderophores (Miethke & Marahiel, 2007; Wandersman & Delepelaire, 2004). *Y. pestis*, the etiologic agent of bubonic plague (defined by the swollen LNs or buboes characteristic of the disease), has 8 inorganic iron and 2 heme/hemoprotein transport systems. In addition to this, *Y. pestis* produces the siderophore yersiniabactin, which is derived from a mixed PKS/NRPS system encoded on the high pathogenicity island (HPI) (Crosa & Walsh, 2002). Interestingly, mutant *Y. pestis* strains deficient in yersiniabactin production or lacking its cognate receptor (Psn) exhibit a dramatic loss of virulence in mice infected subcutaneously (Bearden *et al.*, 1997). *S. pneumoniae* encodes multiple iron transporters that are important for virulence (Brown *et al.*, 2002) and a cell-surface lactoferrin receptor (Hakansson *et al.*, 2001). *S. pyogenes* utilises hemoprotein binding and transport proteins (Lei *et al.*, 2002b) and a multi-metal transport system (Janulczyk *et al.*, 2003). *S. mutans* transports ferrous iron and produces a Mn/Fe uptake system that is important *in vivo* (Evans *et al.*, 1986). Recently, a siderophore-iron (III) transport gene was identified in *S. pyogenes*, *S. agalactiae* and *S. pneumoniae* (Clancy *et al.*, 2006; Hanks *et al.*, 2005; Pramanik & Braun, 2006), but siderophore biosynthesis has not been previously identified in any streptococci (Eichenbaum *et al.*, 1996).

3.1.4 Siderophores

Siderophores are mostly small non-ribosomal peptide molecules, which contain side chains and functional groups that can provide a high-affinity set of ligands for the coordination of ferric ions (Crosa & Walsh, 2002). There are 3 main types of iron-coordinating functional groups in siderophores (Crosa & Walsh, 2002). Anguibactin, produced by the bacterial fish pathogen *Vibrio anguillarum*, has N-hydroxy amino acid side chains, with the oxygen atom as one of the ligands for Fe^{3+}. Adjacent hydroxyls of catechol rings, almost always derived from DHB, are represented in anguibactin and enterobactin (*E. coli*). Enterobactin, which is the strongest iron chelator ever found, has 3 catecholic side chains that can provide a full hexadentate ligation set to ferric ion (Miethke & Marahiel, 2007). A variant on this theme uses salicylate instead of DHB, leading to phenolic moieties as iron ligands. Yersiniabactin and pyochelin (*Pseudomonas aeruginosa*) feature this class of coordinating group and also the 3rd type, namely the nitrogen atoms of 5-membered thiazoline and oxazoline rings (e.g. vibriobactin, *Vibrio cholerae*). Provided that the siderophore contains 6 donor atoms, a 1:1 complex with Fe^{3+} is formed. If there are less than 6 donor atoms provided by the ligand, the vacancies may be occupied by alternative oxygen donors such as water molecules. Alternatively, siderophore complexes with higher stoichiometry may be built up. Pyochelin, for example, can form both 1:1 and 2:1 siderophore: Fe^{3+} complexes (Tseng *et al.*, 2006), whilst cepabactin forms 3:1 complexes (Klumpp *et al.*, 2005; Miethke & Marahiel, 2007).

Siderophores are produced by bacteria during periods of iron starvation and secreted into the extracellular environment to scavenge iron. Very few systems for siderophore secretion have been elucidated. However, since passive diffusion of siderophore molecules is unlikely to occur, dedicated exporters or transporters involved in the efflux of multiple substrates are likely to secrete the siderophore into the extracellular milieu. Siderophore exporters described so far belong to the major facilitator superfamily (MFS)

(Allard *et al.*, 2006; Brickman & Armstrong, 2005; Furrer *et al.*, 2002; Page *et al.*, 2003; Tanabe *et al.*, 2006) and the ATP-binding cassette (ABC) family (Zhu *et al.*, 1998). A resistance, nodulation and cell division (RND) superfamily is involved in antibiotic efflux and has been suggested but not yet proven to be involved in pyoverdine efflux (*P. aeruginosa*) (Li *et al.*, 1995; Miethke & Marahiel, 2007). Upon secretion, extracellular iron-complexes are formed with stabilities ranging over about 30 orders of magnitude for different siderophores. Next, either the iron charged siderophore is taken up by a specific ABC transporter (Chenault & Earhart, 1991; Fetherston *et al.*, 1999; Hanks *et al.*, 2005; Koster & Braun, 1990; Sebulsky & Heinrichs, 2001; Shea & McIntosh, 1991; Wyckoff *et al.*, 1999) or siderophore-bound Fe^{3+} undergoes reduction to Fe^{2+} and is taken up as a single ion through an Fe^{2+}–specific transporter (Miethke & Marahiel, 2007). Systems for extracytosolic reduction utilise free extracellular or membrane-bound reductases but are not common in prokaryotes (Barchini & Cowart, 1996; Miethke & Marahiel, 2007; Schroder *et al.*, 2003; Vartivarian & Cowart, 1999). ABC importers generally require an extracytoplasmic substrate binding protein (periplasmic binding protein or lipoprotein) to scavenge the ferric-siderophore and deliver it to the membrane channel of the ABC transporter (Miethke & Marahiel, 2007). For Gram-negative bacteria, outer membrane receptors are also required to transport the ferric-siderophore across this first barrier (Buchanan *et al.*, 1999; Cobessi *et al.*, 2005; Locher *et al.*, 1998). Bacteria often encode multiple Fe-siderophore importers to allow the use of exogenously derived siderophores as iron sources (Miethke & Marahiel, 2007).

If not already released extracytoplasmatically, the iron has to be removed from the ferric-siderophore complex in the cytosol; a process mediated by intracellular ferric-siderophore reductases or hydrolases (Miethke & Marahiel, 2007). Most Fe-siderophores are supposed to be discharged through the indirect activity of established reductases that are not specific for ferrisiderophore substrates (Halle & Meyer, 1992; Miethke & Marahiel, 2007). Soluble

flavin reductases catalyse the reduction of flavin-based substrates such as flavin mononucleotide (FMN), FAD and riboflavin by using NAD(P)H as an electron donor (Pierre *et al.*, 2002). The reduced flavins are released into solution to serve as reducing agents for a variety of compounds including Fe(III)-chelates. Reduction of Fe^{3+} to Fe^{2+} is followed by spontaneous release or competitive sequestration of the reduced species and may allow recycling of the siderophore. Hydrolytic release is less common and because of damage to the siderophore molecule is likely to be more cost-intensive (Miethke & Marahiel, 2007). The ferric enterobactin esterase (Fes) of *E. coli* does however prevent the potentially toxic effects of intracellular enterobactin accumulation (Furrer *et al.*, 2002).

Siderophore production is tightly regulated to protect against the toxic effects of too much iron. In bacteria, gene regulation of siderophore utilisation and iron homeostasis is generally mediated at the transcriptional level by the ferric uptake repressor Fur (Gram-negative bacteria) or the diphtheria toxin regulator DtxR (Gram-positive bacteria) (Hantke, 2001). DtxR-like proteins also regulate manganese transport in bacteria that regulate iron homeostasis by Fur (Miethke & Marahiel, 2007). Various activators can also contribute to the regulation of siderophore utilisation through the direct or indirect sensing of extra- or intracellularly present ferric-siderophores. These include alternative sigma factors, 2 component sensory transduction systems and AraC-type regulators (Miethke & Marahiel, 2007).

3.2 Aims

Genome sequencing and comparative analysis of *S. equi* and its close ancestral relative *S. zooepidemicus* has enabled the identification of possible virulence determinants in the former that could contribute to its increased pathogenicity (Chapter 2). ICE*Se2* was identified by this process and detected in all strains of *S. equi* examined. The acquisition

of ICE*Se2* by *S. equi* was hypothesised to have been an important event in the speciation of this subspecies owing to its apparent absence from the *S. zooepidemicus* population.

The overall aims of the work presented in this chapter were to determine the function of the NRPS encoded by ICE*Se2* in *S. equi* and to establish the contribution of individually encoded proteins to the biosynthesis, transport, regulation and virulence (in mice) of the NRPS product. These aims were achieved using the following specific objectives:-

1. PCR and Southern blot analysis to establish the prevalence and genome location of ICE*Se2* across a diverse selection of *S. equi* and *S. zooepidemicus* strains.
2. Bioinformatic analysis to assess the similarity of the ICE*Se2* NRPS to other characterised NRPS systems.
3. Structure-based bioinformatic analysis and biochemical analysis to determine the structure of the ICE*Se2* NRPS product.
4. Allelic-replacement mutagenesis coupled with *in vitro* growth analyses to determine the role of the ICE*Se2* NRPS in *S. equi* cation (and more specifically Fe^{3+}) incorporation.
5. Allelic-replacement mutagenesis coupled with *in vitro* growth in chemically defined media to determine the biochemical requirements of the ICE*Se2* NRPS.
6. Allelic-replacement mutagenesis coupled with *in vitro* growth analyses to identify the mechanisms for transport of the ICE*Se2* NRPS product.
7. Allelic-replacement mutagenesis coupled with quantitative reverse transcription-PCR to determine the transcriptional regulation of the ICE*Se2* NRPS *eqbE* gene by the proposed EqbA repressor.
8. Electrophoretic mobility shift assays (EMSAs) to assess the ability of recombinant EqbA to bind to the promoter region of the *eqbB* gene (the 1st gene of the proposed

NRPS operon) and to determine the influence of some divalent cations (and Fe^{3+}) on such binding activity.

9. Allelic-replacement mutagenesis to establish the contribution of the *eqbE* putative NRPS gene to the virulence of *S. equi* in mice.

3.3 Methods

3.3.1 Bacterial strains, media and growth conditions

*Se*4047 was isolated in 1990 from the submandibular abscess of a New Forest pony. Strains were cultured in THB (Sigma) or on Todd Hewitt agar (THA) (Sigma) at 37°C with 5% CO_2 unless otherwise stated. Chemically defined media (CDM) for Group A Streptococci (SAFC Biosciences, Appendix Table A.3) was used to produce conditioned supernatant for siderophore assays and to identify the additional biochemical requirements of *eqb* NRPS product production on supplementation with 10 µM salicylate (Sigma) (CDMs). To measure the influence of free cation concentration on the growth of wild-type and mutant *S. equi* strains, THA was supplemented with 2 mM nitrilotriacetic acid (NTA, Sigma).

3.3.2 Identification of *Streptococcus equi* restricted genes

ACT (Carver *et al.*, 2005) was used to view TBLASTX alignments of the publicly available sequence data from the *S. equi* and *S. zooepidemicus* genome sequencing projects (www.sanger.ac.uk accessed 10.03.09). Comparisons of predicted CDSs with Uniprot were performed using BLASTP and FASTA. Southern blot and PCR analyses were used to determine the genome location, circularisation and prevalence of ICE*Se2* across 18 strains of *S. equi* and 73 strains of *S. zooepidemicus* (Table 3.1, Appendix Table A.4). Diverse strains were selected based on their SeM allele (Kelly *et al.*, 2006) and RNA intergenic spacer/M protein sub-type (Newton *et al.*, 2008), respectively (Appendix Table

A.4). Taq DNA polymerase (Sigma) was used for PCR according to manufacturer's instructions.

Table 3.1 Oligonucleotide primers used for PCR to screen diverse *S. equi* and *S. zooepidemicus* strains for the presence of *eqbE*, to amplify across flanking regions of ICE*Se2*, to generate DNA probes for Southern blot analysis and electrophoretic mobility shift assays, to clone plasmid constructs for gene expression or the production of allelic replacement *S. equi* deletion mutants and to perform quantitative RT-PCR analysis of *eqbE*/*gyrA* gene expression

Primer	Sequence (5'-3')	Purpose
ZM236	TTTTTTCTTCTTCCCACTGGC	Left flank of ICE*Se2* (498 bp) and circular form of ICE*Se2* (~537 bp)
ZM343	TTGGGATGACCATGGGATAC	Left flank of ICE*Se2* (498 bp)
ZM235	ATTGGGAACACCTTGCAAGG	Right flank of ICE*Se2* (1590 bp) and circular form of ICE*Se2* (~537 bp)
ZM344	TGTTGGTGTTCCTGGGATTC	Right flank of ICE*Se2* (1590 bp)
ZM233	CGTGATTTGAGAGTAGGGAC	*eqbN* probe for Southern blots (317 bp)
ZM234	TGTAACGAAGTCGCTACTGC	
EqbE f	AAGATATAGCAGCATCGTATCG	Sense primer QRT-PCR of *eqbE* (130 bp)
EqbE r	TCTAAATCTCTATTAAATAGCGGTATATTG	Anti-sense primer
GyrA f	AAGGCGGGATTCCTAAAATC	Sense primer QRT-PCR of *gyrA* (143 bp)
GyrA r	GATAAGTAAGCCCTCTAAAATGTG	Anti-sense primer
2f	GGGTTGCCATGCATATCTTG	*eqbE* screen (548 bp)
2r	TCCGGCTGTTTCCTTAATGG	
ZM94	GAGGTCGACGTCACAACAAGACTCTTCCC	*eqbE* deletion (amino acids 4 to 2017)
ZM95	GACGATATCATTATAGAAGGGAGTTTATGATG	
ZM96	GACGATATCAAGTTCCATTCCAAAATCCTCC	
ZM97	GACGAATTCTCGGAATCGCTGAAGGATTG	
ZM123	GACGAATTCACTTTTACAACCGGACAGTTC	*eqbA* deletion (amino acids 2 to 124)
ZM124	GACGATATCATAAACTCATCTAACTTATCCCTTC	
ZM125	GACGATATCCAAGATATAAGAGAAAGAACATGC	
ZM126	GACGTCGACCCATTCGGATTTGGTAACCC	
ZM199	GACGCGGAGGAATTCTTGCTGTAG	*eqbHIJ* deletion (amino acids 3-*eqbH* to 419-*eqbJ*)
ZM198	GACGACGATATCGCTCATTATTATTCTCCTCTGTC	
ZM178	GACGACGATATCGAAGGTATGCTTGCAGTGTC	
ZM179	GACGACGTCGACAACATCGCAAGAGCCATCTC	

Table 3.1 continued

Primer	Sequence (5'-3')	Purpose
ZM162	GACGAATTCGGGTCTTAGAACTTTAGAGG	*eqbKL* deletion (amino acids 3-*eqbK* to 572-*eqbL*)
ZM163	GACGATATCGCTCATGTAACTTCTCCACC	
ZM164	GACGATATCAGGGGTAAAACAGAGGACTG	
ZM165	GGGGTCGACCTGGCATACAAATAACGTCTCC	
ZM180	GACGACGAATTCTTACCTCAGCTGCAAGAAGC	*ftsB* deletion (amino acids 23 to 304)
ZM181	GACGACGATATCACTTGAGCAGGCGACTAATG	
ZM182	GACGACGATATCATCCTCTCAACTGGTGCAAG	
ZM183	GACGACGTCGACAGCCTCATTTGAGTGTAGCC	
ZM119	GACGAATTCCATGCTAAGGAGGGTTTTCC	*fur* deletion (amino acids 3 to 155)
ZM120	GACGATATCTCCAAGCTTACAACCTCCTC	
ZM121	GACGATATCCACAGACTATTGAGAGCTGG	
ZM122	GACGTCGACTGCCAGTTCAGATGTTGACC	
ZM329	GACGACGGATCCTTATGAATAAAATATATCATA AGG	Construction of pGEX-*eqbA*
ZM330	GACGACGAATTCCATGTTCTTTCTCTTATATCTT G	
BIOZM 360	AAGATATAAGAGAAAGAACATGC	Amplification of 5' biotin end-labelled target DNA for EMSA
ZM361	TCCATCTGATTATTAATATACTAC	Used with BIOZM360 to amplify 227 bp P_{eqb} target DNA A for EMSA
ZM362	TTAATCCTTCTAACATAAAAAGTC	Used with BIOZM360 to amplify 165 bp target DNA B for EMSA
ZM452	CTAAAAGGCTTGAAGAAACTAAG	Used with BIOZM360 to amplify 103 bp target DNA C for EMSA
ZM446	GACGACGAATTCGTAGTATAACAATAATAAGAA TGTTG	pA*eqbA* for cloning pGpA*eqbA*
ZM447	GACGACGACGTCGACTCTTTCTCTTATATCTTGT CTTC	
ZM442	GACGACGACCCCGGGCAAGATATAAGAGAAAG AACATGC	pB for recombinant PCR
ZM443	CTTTCTTTTATATTGTTAAGTTCCATCAAAACAC CTTCCATCTGATTATTAATATAC	
ZM444	GTATATTAATAATCAGATGGAAGGTGTTTTGAT GGAACTTAACAATATAAAAGAAAG	*eqbE*-Nterm for recombinant PCR
ZM445	GTTCTAAGCATGGGATGTCTG	

Restriction sites are underlined.

For Southern blot analysis 2 to 4 μg genomic DNA was digested with NdeI (NEB) overnight. NdeI restriction sites flank an 8.2 kb DNA fragment containing part of the *eqbN* gene, the *eqbM* gene, the C-terminal and downstream regions of ICE*Se*2. Digested DNA was separated by gel electrophoresis and transferred to a positively charged nylon membrane (Roche Applied Science) by capillary action following depurination (0.25 M HCl, 20 min), denaturation (0.5 M NaOH, 1.5 M NaCl, 2 x 15 min) and neutralisation (1 M Tris-HCl pH 7.4, 1.5 M NaCl, 2 x 15 min) treatments. The capillary transfer was performed overnight using standard apparatus and 20 x SSC buffer (3 M NaCl, 0.3 M sodium citrate, pH 7.0). Transferred DNA was fixed to the nylon membrane by UV transillumination (3 min). The membrane was then pre-equilibrated with Dig Easy Hyb (Roche Applied Science) (30 min, 40°C) and probed with DIG (digoxigenin)-labelled *eqbN* probe (denatured at 95°C for 5 min), diluted 1/500 in Dig Easy Hyb (40°C overnight). DIG-labelled *eqbN* probe was prepared using the PCR DIG Probe Synthesis Kit (Roche Applied Science) according to the manufacturer's instructions. Unbound probe was washed from the membrane with 2 x SSC, 0.1% SDS (2 x 5 min, RT) followed by 0.5 x SSC, 0.1% SDS (2 x 15 min, 42°C). Bound probe was detected using the DIG Luminescent Kit for Nucleic Acids (Roche Applied Science) using anti-DIG-AP (alkaline phosphatase) conjugate and chemiluminescent detection according to the manufacturer's instructions.

3.3.3 Allelic replacement

Internal gene deletions were introduced into *Se*4047 through an allelic replacement strategy previously described for the production of a Δ*prtM* mutant (Hamilton *et al.*, 2006). First, a copy of the *Se*4047 gene (with flanking regions) containing a deletion was introduced into the pGhost9 plasmid by 3-way ligation. Phusion DNA polymerase (NEB) was used to generate DNA fragments by PCR, which were purified with Qiagen spin

columns. Table 3.1 shows the primers used and details internal deletions in the construction of plasmids pG*eqbA*Δ, pG*eqbE*Δ, pG*eqbHIJ*Δ, pG*eqbKL*Δ and pG*ftsB*Δ. Cloning was performed in *E. coli* TG1repA+ cells, plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen) and sequencing with an ABI3100 DNA sequencer was performed to check for errors (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems). In all cases, kits were used according to the manufacturer's instructions. Plasmid constructs (1 – 5 μg) were transformed into *Se*4047 by electroporation (Bio-Rad Gene Pulser electroporater with pulse settings of 2.5 kV cm–1, 200 Ω, and 25 μF) following growth of cells (10 ml) to early log phase in THB supplemented with hyaluronidase (30 μg ml^{-1}), washing (3 x 1 ml) and resuspension in ice-cold 0.5 M sucrose solution (100 μl). Transformants were selected by growth at 28°C (for 2 to 3 days) on THA supplemented with 0.5 μgml^{-1} erythromycin following a 2 to 3 h recovery time. pGhost9 is able to replicate at 28°C, but can only be maintained at 37°C by integrating into the bacterial chromosome. 2 rounds of homologous recombination were used to replace the wild-type gene with its respective deletion allele. The first recombination event, leading to integration of the plasmid construct, was achieved following growth at 37°C and the second recombination event, leading to the excision of the plasmid construct was performed at 28°C. PCR was used to identify successful integration events and to screen erythromycin sensitive colonies for the presence of a gene deletion. Sequencing was also used to confirm that a deletion had occurred and that no sequence errors were present in flanking regions.

The Δ*eqbAE* double deletion mutant was generated by transformation of *S. equi* strain Δ*eqbE* with the pG*eqbA*Δ plasmid followed by recombination and selection as described previously. Similarly, for other multiple deletion strains, *eqbA* was always the last gene to be deleted.

3.3.4 RNA isolation

Total RNA was extracted from bacteria using RNAprotect, RNeasy and DNase Kits according to the manufacturer's instructions (Qiagen). Bacteria grown to mid-log phase were pelleted by centrifugation for 10 min at 5000 g. The pellet was resuspended in 200 µl of TE buffer (10 mM Tris-Cl, 1 mM EDTA; pH 8.0) containing 3 mg lysozyme (Sigma) and 500 U mutanolysin (Sigma). After incubation at room temperature for 45 min with repeated vortexing, 700 µl of RLT buffer containing β-mercaptoethanol was added, and the tube was vortexed. The mixture was transferred to a 2 ml reaction tube containing 0.05 g of 100 µm-diameter acid-washed glass beads (Sigma) and vortexed for 5 min. The mixture was then centrifuged and total RNA was extracted from the supernatant using an RNeasy Midi Kit. 5 µg of recovered RNA was treated with DNase to remove any contaminating DNA followed by clean up on an RNeasy mini column. The quantity and purity of RNA was determined by optical density measurements at 260 and 280 nm using a NanoDrop^R ND1000 spectrophotometer (NanoDrop Technologies).

3.3.5 Reverse transcription and quantitative real-time PCR

Transcriptional regulation of the NRPS gene, *eqbE* by the putative repressor, EqbA was determined by quantification of *eqbE* transcript levels in the Δ*eqbA* and parent *Se*4047 strains grown to mid-log phase in THB. A quantitative two-step reverse transcription PCR procedure was used to calculate levels of *eqbE* gene transcription relative to the housekeeping gene *gyrA*. Reverse transcription (RT) was performed using the Verso cDNA kit (Abgene). The RT reaction mixture (20 µl) contained 96 ng total RNA, 2 µM gene-specific anti-sense primer (Table 3.1), 500 µM dNTP mix, 16 U RNase inhibitor (RNASIN, Promega), 1x cDNA synthesis buffer, 1 µl RT enhancer and 1 µl Verso enzyme mix. RT was performed at 50°C for 30 min and terminated by heating to 95°C for 2 min. QPCR was performed with a Techne Quantica instrument and data analysed using Quansoft software (Techne). For the QPCR, 6 µl RT reaction mixture diluted 1/1000 was

mixed with 0.3 μM forward and reverse primers (Table 3.1), 1x ABsolute QPCR SYBR green mix (Abgene) and 40 nM ROX in a total volume of 20 μl and subjected to thermocycling at 95°C for 15 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Dissociation curves were analysed, following a final ramp step from 60°C to 90°C with reads at 0.5°C increments, to rule out non-specific amplification. No-template negative controls were included and reverse transcriptase negative controls to confirm the absence of contaminating DNA from RNA samples. Standard curves (Cp vs log gene copy number) with an efficiency of 1.9 ($R^2 = 0.999$) were generated from gDNA for each target gene and used to calculate gene copy number in cDNA samples generated from 3 independent RNA extractions. *eqbE* gene copy number was normalised to *gyrA* reference gene copy number to correct for differences in the amount of starting material. Data was expressed as fold increase in normalised *eqbE* transcript level in the *ΔeqbA* mutant strain relative to the wild-type *Se*4047 strain.

3.3.6 Streptonigrin sensitivity

The minimum inhibitory concentration (MIC) of streptonigrin and erythromycin was determined following overnight incubation of bacterial strains in a standard 96-well microtiter plate assay. 100 μl of re-suspended colony material corresponding to 10^4 colony forming units (CFU) ml^{-1} in THB, CDMs or conditioned THB/CDMs was added to 100 μl THB supplemented with streptonigrin (concentration ranging from 2 μM to 0.5 nM through two-fold serial dilutions) in duplicate wells of a microtiter plate. Sensitivity to erythromycin was determined using the same procedure (0.5 μg ml^{-1} to 0.5 ng ml^{-1} erythromycin). Data is presented as the mode of at least 3 independent experiments. 2-sample Wilcoxon rank-sum (Mann-Whitney) tests were used to measure the significance of differences in streptonigrin sensitivity between strains or conditioned media samples (performed by Richard Newton, AHT, using Stata 9.2 software (StataCorp LP)).

3.3.7 $^{55}FeCl_3$ accumulation

Wild-type and mutant *S. equi* strains were depleted of iron by growth overnight on THA supplemented with 4 mM NTA. Colony material from these plates was re-suspended in PBS with the volume adjusted to give an optical density at 600 nm of 1.0. 0.5 ml of this bacterial suspension was added to 0.5 ml THB supplemented with 2.0 μCi ml^{-1} $^{55}FeCl_3$ (Perkin Elmer) and incubated at 37°C with 5% CO_2. After 3 hr, cell growth was monitored through measurement of optical density and a volume of cells equivalent to 6 x 10^7 CFU (optical density of 0.3 ≡ 2 x 10^8 CFU ml^{-1}) was collected in triplicate wells of a 96-well UniFilter^R GF/CR (Perkin Elmer) with a Packard FilterMate^{TM} Harvester. Cells were washed on the filter with 10 mM NTA and then methanol. The filter was dried for 1 hr at room temperature prior to the addition of 20 μl Microscint 20 (Packard). Bacterial iron accumulation was measured by liquid scintillation counting on a TopCount NXT^{TM} microplate counter (Packard). 2-sample Wilcoxon rank-sum (Mann-Whitney) tests were used to measure the significance of differences in mean $^{55}FeCl_3$ accumulation between the Δ*eqbA* strain and Δ*eqbAE,* Δ*eqbHIJA* and wild-type strains (performed by Richard Newton, AHT, using Stata 9.2 software (StataCorp LP)).

A modification of this method was used to analyse the accumulation of $^{55}FeCl_3$ in the *S. equi* mutant strain *ΔeqbAE* when cross-fed with filter sterilised culture supernatant from wild-type, Δ*eqbAE* or Δ*eqbA* strains grown to stationary phase overnight. Bacteria were resuspended in 1 ml 2X THB supplemented with 2.0 μCi ml^{-1} $^{55}FeCl_3$ to an optical density at 600 nm of 0.2 and mixed with an equal volume of filter sterilised culture supernatant. Cells were incubated at 37°C with 5% CO_2 until they reached an optical density at 600 nm of 0.5 and then a volume of cells equivalent to 1 x 10^8 CFU was collected in triplicate wells of a 96-well UniFilter^R GF/CR and processed as described above.

3.3.8 LC-MS analyses of culture supernatants

Comparative metabolic profiling of the Δ*eqbA*, Δ*eqbAE* and Δ*eqbHIJA* strains was performed using LC-MS (liquid chromatography-mass spectrometry) by Lijiang Song under the supervision of Greg Challis (Warwick University). Culture supernatants (50 ml) were extracted with ethyl acetate (3 x 50 ml) and both the combined organic extract and the residual culture supernatant were concentrated to dryness and re-dissolved in 20% acetonitrile / 0.1% formic acid in water (1 ml). Each sample was split into two equal portions and ferric chloride was added to one portion of each sample to a final concentration of 5 mM. 50 µl of each sample were analysed on an Agilent Zorbax C-18 reverse phase column (150 x 4.6 mm) connected to an Agilent 1100 high performance liquid chromatography (HPLC) instrument. The outflow was transferred via a splitter (90% to waste, 10% to mass spectrometer) to a Bruker HCT+ mass spectrometer, equipped with an electrospray source, operating in positive ion mode. The column was eluted as follows (mobile phase A: water with 0.1% formic acid, mobile phase B: acetonitrile with 0.1% formic acid). 100% mobile phase A for 5 min; a gradient from 100% mobile phase A to 100% mobile phase B over 25 min; 100% mobile phase B for 5 min; a gradient from 100% mobile phase B to 100% mobile phase A over 5 min; 100% mobile phase A for 5 min.

3.3.9 Chrome azurol S assay

Iron-chelating activity in supernatant samples was monitored by the chrome azurol S (CAS) assay and CAS shuttle assay (Schwyn & Neilands, 1987). Bacteria were grown overnight in 10 ml CDMs to stationary phase and then filter sterilised through a 0.22 µm filter (Millipore) to recover supernatant. 0.5 ml supernatant was mixed with an equal volume of CAS assay solution (or CAS shuttle assay solution) and change in absorbance at

630 nm was measured over time. Desferoxamine was included as a standard. CAS assay solution (made up to 100 ml in H_2O) contained 6 ml 10 mM hexadecyltrimethylammonium (HDTMA), 1.5 ml iron solution (1 mM $FeCl_3$, 10 mM HCl), 7.5 ml 2 mM CAS solution (0.06 g in 50 ml H_2O), 4.3 g anhydrous piperazine dissolved in H_2O, 6.25 ml 12 M HCl. CAS shuttle assay solution also contained 4mM 5-sulfosalicylic acid. This reagent is intended to act as a shuttle to aid the removal of iron from the CAS-HDTMA-iron complex by low affinity iron chelators. All reagents were from Sigma unless otherwise indicated.

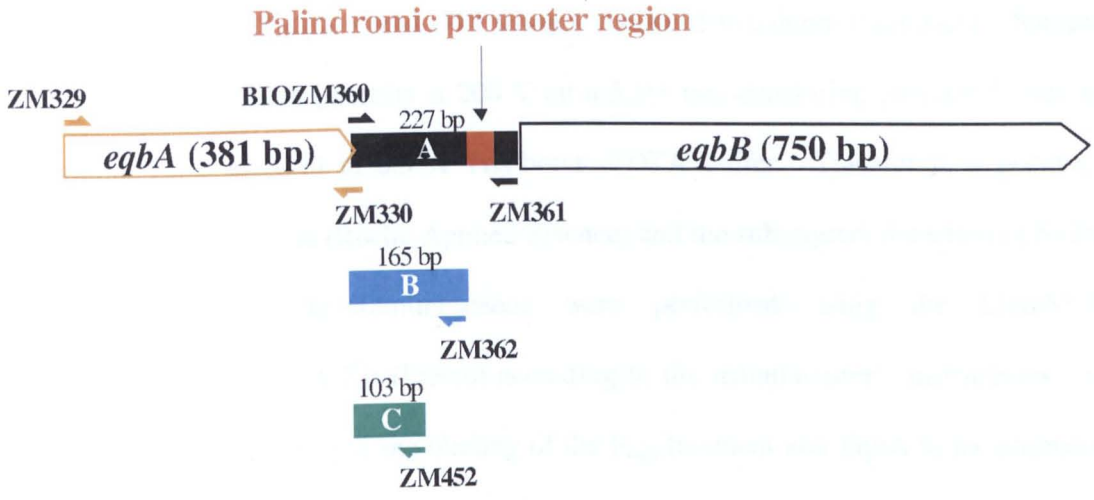
Iron-chelating activity was also assessed using CAS agar overlayed with bacteria resuspended in THA in 15 ml polystyrene tubes. 100 ml of CAS agar contained 1.2 g agar, 5 ml CAS solution (0.06 g CAS in 50 ml H_2O), 1 ml iron (III) solution (1 mM $FeCl_3$ in 10 mM HCl) and 4 ml HDTMA (0.07 g in 40 ml H_2O). Samples were cultured at 37°C with 5% CO_2 for 5 days. *S. aureus* was included as a positive control.

3.3.10 Overexpression and purification of EqbA

The *S. equi eqbA* gene was amplified by PCR from *Se*4047 chromosomal DNA with Phusion polymerase (NEB), primer ZM329 and primer ZM330 (Table 3.1, Figure 3.1). The PCR product was cloned into the BamHI/EcoRI sites of the pGEX-3X vector (GE Healthcare). The resulting construct, pGEX-*eqbA*, contains an N-terminal fusion of *eqbA* to a glutathione S-transferase (GST) tag driven by a tac promoter. Subsequent expression and purification of EqbA (126 amino acids) was performed by Carl Robinson (AHT). DH10B *E. coli* cells harbouring the plasmid pGEX-*eqbA* were grown at 37°C with shaking (220 rpm) in 2 x YT containing 50 µg/ml ampicillin. Once the cells reached an optical density at 600 nm of 0.6, 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the medium and the culture was incubated for 4 h at 28°C with shaking (220 rpm). Cells were then harvested and lysed by sonication, and the GST-EqbA fusion was purified over

glutathione sepharose 4B beads according to the supplier's protocol (GE Healthcare). Recombinant EqbA was cleaved from the beads by Factor Xa protease cleavage in 50 mM Tris-HCl (pH 7.5), 150 mm NaCl, and 1 mM $CaCl_2$ according to the supplier's protocol (GE Healthcare). The purified EqbA protein appeared homogenous by SDS-polyacrylamide gel electrophoresis and coomassie blue staining.

Figure 3.1 Diagram showing primer binding sites used for the amplification of *eqbA* and electrophoretic mobility shift assay DNA fragments by PCR



The *eqbA* gene was amplified using primers ZM329 and ZM330 for cloning and expression using pGEX-3X. 3 DNA fragments (A, B and C) upstream of the *eqbB* gene were amplified by PCR using the primers indicated in order to assess their binding to recombinant EqbA using electrophoretic mobility shift assays (EMSAs). Only fragment A contained the palindromic promoter region highlighted in red.

3.3.11 Electrophoretic mobility shift assay

5' biotin end-labelled DNA fragments A, B and C containing the upstream region of the *eqb* operon were amplified from the *Se*4047 chromosome using the primers indicated in Table 3.1 and Figure 3.1. Fragment A contains the -237 to -11 bp region upstream of the *eqb* operon (P_{eqb}), whilst fragments B and C contain regions -237 to -73 bp and -237 to -135 bp, respectively and lack the DNA palindrome of P_{eqb} (-73 to -38 bp). The PCR products were purified from a 1.5% agarose gel using a gel extraction kit (Qiagen).

Binding reactions (20 µl) were carried out at room temperature for 20 min in 22.5 mM Tris-HCl (pH 7.5), 67.5 mM KCl, 5 mM $MgCl_2$, 1.45 mM dithiothreiotol (DTT), 5% glycerol and 1 µg dIdC. 20 fmol biotin-labelled target-DNA were mixed with 300 pmol EqbA with or without the addition of 125 µM freshly prepared $FeSO_4$, $ZnSO_4$, $MnCl_2$, $CuSO_4$ or $FeC_6H_5O_7$. To assess the specificity of rEqbA binding to P_{eqb}, 3.8 pmol unlabelled P_{eqb} was added 5 min prior to the addition of biotin-labelled P_{eqb}. To remove divalent metal-ion from the purified EqbA, the protein was incubated with 2 mM EDTA for 2.5 h with mixing at room temperature and then buffer exchanged against 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 1 mM DTT using a vivaspin500 column (Sartorius). Samples were analysed by electrophoresis at 200 V on a 5.5% non-denaturing polyacrylamide gel containing 2.5% glycerol in 0.5 X Tris-borate-EDTA buffer. Transfer to a positively charged nylon membrane (Roche Applied Science) and the subsequent detection of biotin-labelled DNA by chemiluminescence were performed using the LightShift^R Chemiluminescent EMSA Kit (Pierce) according to the manufacturer's instructions. Kit controls were used to analyse the binding of the P_{eqb} fragment and EqbA to an alternative protein extract and target DNA respectively.

3.3.12 Experimental infection of mice

2 repeat experiments were performed to assess the virulence of the Δ*eqbE* mutant in comparison to the wild-type strain of *S. equi* in mice. Mice were challenged intranasally as previously described with some modifications (Chanter *et al.*, 1995; Flock *et al.*, 2004). To infect mice, bacteria were inoculated into THB supplemented with 10% foetal calf serum (Sigma) (THBfcs) and grown overnight. The culture was then diluted 20 times into THBfcs and grown to an optical density of 0.3 at 600nm. 20 7 to 8 week old female BalbC mice were challenged by intranasal administration of $1x10^4$ CFU in a 10µl volume with either the parent *Se*4047 or Δ*eqbE S. equi* strain. In order to assist complete delivery of the

innoculum into the nasal cavity and therefore improve infectivity rate, challenge was performed under sedation with Ketaset™ (Genus) administered intra-muscularly at 100 $mgkg^{-1}$ bodyweight. Mice were examined daily and clinical signs of disease such as weight loss and general demeanour were compared with a group of 5 unchallenged controls over a period of 7 days. An overall clinical observation score was assigned based on the sum of scores allocated for 9 categories: normal (0), slight piloerection (1), piloerection (3), marked sneezing (3), quietness (3), hunching (5), rapid breathing (5), ear imbalance (10) and immobility (10). Nasal shedding was also determined daily by gently touching the nostrils to 5% blood agar plates (bioMerieux). A sterile spreader was used to distribute the inoculum across the plate and bacterial growth following overnight incubation was scored (experiment one) or the number of CFU were counted (experiment 2). Bacterial growth was scored on a 5-point scale from 0 to 4, where 0 means no growth, 1 means 1 to 10 CFU, 2 means 11 to 100 CFU, 3 means > 100 CFU and 4 means confluent growth. Any mice losing > 15% of their body weight or showing signs of suffering were considered to have reached the clinical end point and were euthanased, whilst all other mice were euthanased at the end of the 7 day period. All mice were examined post-mortem for evidence of pneumonia or submandibular LN abscessation and heart swabs were taken to check for septicaemia (experiment 2 only). 2-sample Wilcoxon rank-sum (Mann-Whitney) tests were used to measure the significance of differences in mean cumulative weight loss per day and mean cumulative gain in nasal score per day between the *ΔeqbE* and wild-type infected groups (performed by Richard Newton, AHT, using Stata 9.2 software (StataCorp LP)). Debs Flack, Beverley Walladge, Lyndsey Owen, Andrew Waller and Carl Robinson (AHT) helped with the daily care, health checks and around the clock monitoring of the mice.

3.4 Results

3.4.1 Identification and prevalence of a novel ICE encoding a siderophore-like NRPS in *Streptococcus equi*

Comparison of the *Se*4047 genome and *Sz*H70 genomes using ACT (Carver *et al.*, 2005) led to the identification of a large 63 kb locus in *Se*4047 that was absent from the *S. zooepidemicus* H70 genome, had a significant decrease in G+C content (31% compared to 42% across the whole *Se*4047 genome) and was bordered by 108 bp perfect inverted repeat elements.

BLASTP and FASTA analysis against UniProt determined that the locus contained coding regions with similarity and conserved gene order with CDSs contained within the conjugative transposons CDTn*2* and CDTn*5* in the genome of *C. difficile* strain 630 (Sebaihia *et al.*, 2006) and the plasmid borne Tn*1549* in *Enterococcus spp.* (Garnier *et al.*, 2000) (Figure 3.2). ICE*Se2* also shares some homology with an ICE in *S. pyogenes* MGAS10750 (ICE MGAS10750-RD.2, Figure 3.3) (Beres & Musser, 2007). Consequently I have named this *S. equi* integrative conjugative element, ICE*Se2*, according to the suggested nomenclature (Burrus *et al.*, 2002b). Like Tn*916*, these elements appear modular, with CDSs likely to be associated with conjugation and site-specific recombination located in the left and right extremities respectively. The recombination module of ICE*Se2* encodes a single serine recombinase (284 amino acids) and has homology (64%-43% amino acid sequence identity) to the N terminal regions (first 277-303 amino acids) of similar proteins in ICE MGAS10750-RD.2, CDTn*2* and CDTn*5*, and also TndX (533 amino acids, 43% amino acid sequence identity), which is responsible for the integration and excision of the conjugative transposon Tn*5397* (Wang *et al.*, 2000). The sequence and annotation of ICE*Se2* has been submitted to GenBank/EMBL with the accession number AM909652.

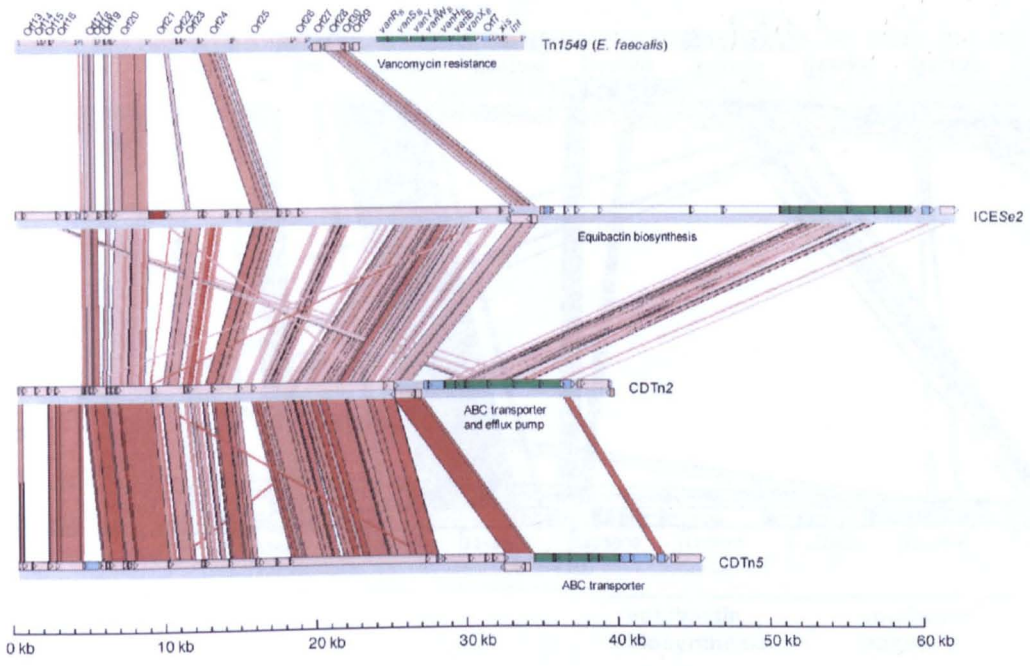


Figure 3.2 Comparison of ICE*Se2* with the conjugative transposons CDTn*2* and CDTn*5* of *C. difficile* Strain 630 (Sebaihia *et al.*, 2006) and Tn*1549* from *E. faecalis* (Garnier *et al.*, 2000)

Figure reproduced with kind permission from Matt Holden (displayed using ACT (Carver *et al.*, 2005)). Grey bars represent the forward and reverse strands of DNA. The red lines between the elements represent protein similarity (TBLASTX). CDSs involved in similar function are given the same colour: pink, conjugal transfer and integration and excision; blue, regulation; white, secondary metabolism; green, transport; red, accessory function. The gene names of Tn*1549* are shown on top. The genes encoding the equibactin NRPS are shown in white and those encoding the ABC transporters associated with the equibactin locus are shown in green.

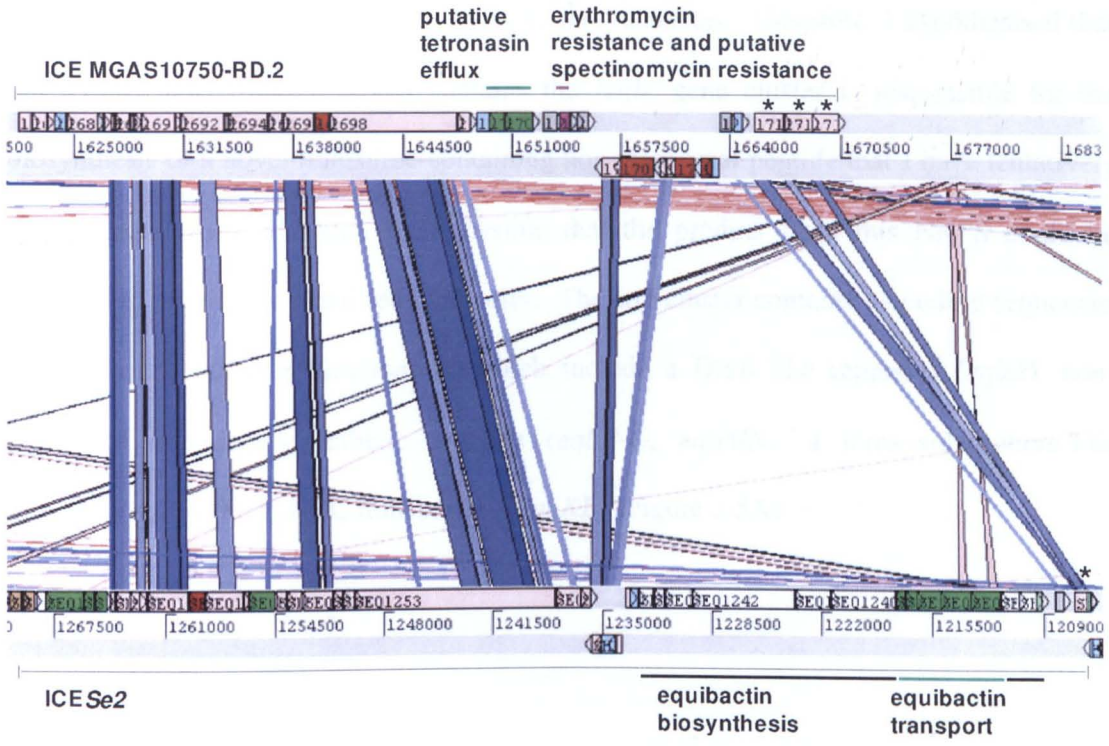


Figure 3.3 Comparison of ICE*Se2* with the MGAS10750-RD.2 ICE of *S. pyogenes* MGAS10750

Alignment displayed using ACT. The *Se*4047 genome region encoding ICE*Se2* is shown on the bottom and the *S. pyogenes* genome region encoding MGAS10750-RD.2 ICE (Beres & Musser, 2007) is shown on the top. Grey bars represent the forward and reverse strands of DNA with numbering in between to indicate genome location. The blue (and red) lines between the elements represent protein similarity (TBLASTX, cut-off score of 75). CDSs involved in similar function are given the same colour: pink, conjugal transfer and integration and excision (* putative serine recombinase); blue, regulation; white, secondary metabolism; green, transport; red, accessory function; majenta, erythromycin and putative spectinomycin resistance; grey, hypothetical and unclassified; brown, pseudo- or partial gene. The MGAS10750-RD.2 ICE carries genes encoding erythromycin resistance, putative tetronasin resistance and putative spectinomycin resistance. ICE*Se2* encodes equibactin NRPS biosynthetic genes and associated transporters.

The 'cargo' region between the conjugation and recombination modules that encodes vancomycin and erythromycin resistance in Tn*1549* and ICE MGAS10750-RD.2, respectively, carries transport genes of unknown function in CDTn*2/5* and CDSs in ICE*Se2* with most overall similarity to the NRPS cluster 1 of *Clostridium kluyveri*, which is proposed to biosynthesise a putative siderophore (Seedorf *et al.*, 2008) (Figure 3.4). Several of the encoded proteins were also similar to the NRPS complex of *Yersinia sp.* that

produces the ferric iron-binding siderophore yersiniabactin. Therefore, I hypothesised that the ICE*Se2* cargo region which contains the '*eqb*' gene cluster is responsible for the biosynthesis of a novel thiazoline-containing non-ribosomal peptide that I have tentatively named 'equibactin' although it is possible that the product from this NRPS could be identical to previously identified molecules. The *eqb* cluster contains 14 coding sequences (*eqbA-N*), the putative functions of which include a DtxR-like repressor (*eqbA*), non-ribosomal peptide biosynthetic proteins (*eqbB-G, eqbMN*), a ferric-siderophore-like importer (*eqbH-J*) and ABC transporters (*eqbKL)* (Figure 3.5A).

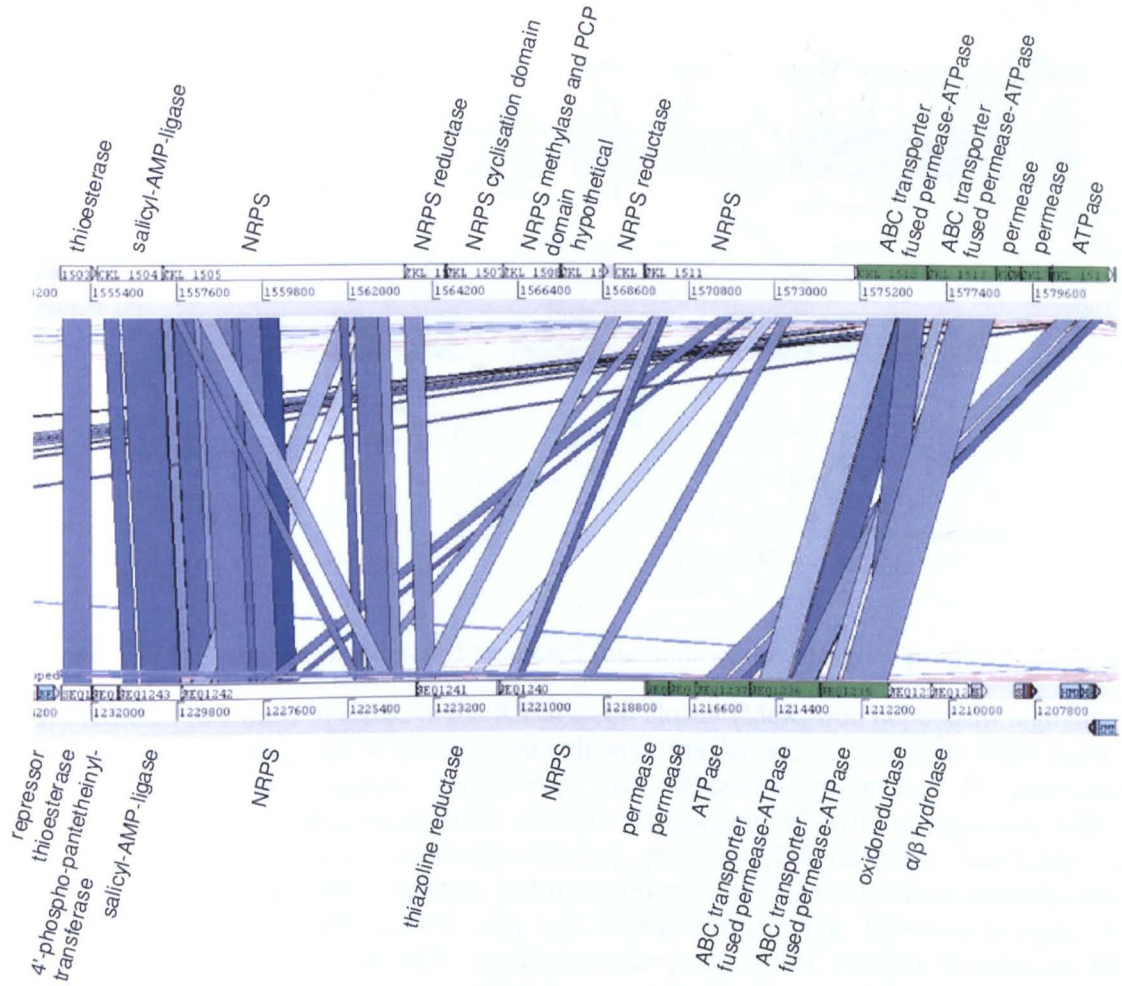


Figure 3.4 Comparison of the ICE*Se2* NRPS cluster with NRPS cluster 1 in *C. kluyveri* strain DSM555

Alignment displayed using ACT. The *C. kluyveri* NRPS cluster 1 (Seedorf *et al.*, 2008) is shown on top and includes 9 CDSs putatively involved in non-ribosomal peptide biosynthesis (CKL_1503 to CKL_1511) and 5 downsteam putative ABC transporter CDSs (CKL_1512 to CKL_1517). The putative functions of 10 CDS immediately downstream of CKL_1517 (not shown) include a multi-antimicrobial extrusion (MATE) family drug/sodium antiporter (CKL_1518), a hypothetical protein (CKL_1519), a glycosyltransferase (CKL_1520), an ATPase (CKL_1521) and a membrane protein containing 12 predicted TMHs (CKL_1522), followed by a PKS gene cluster (CKL_1523 to CKL_1527, cluster 2 (Seedorf *et al.*, 2008)). The ICE*Se2* NRPS cluster is shown on the bottom and includes an upstream regulator (*eqbA*, SEQ1246), 8 CDS involved in non-ribosomal peptide biosynthesis [*eqbB* (SEQ1247) to *eqbG* (SEQ1240), *eqbM* (SEQ1234) to *eqbN* (SEQ1233)] and 5 downstream ABC transporter CDSs [*eqbH* (SEQ1239) to *eqbL* (SEQ1235)]. Grey bars represent the forward and reverse strands of DNA with numbering in between to indicate genome location. The blue lines between the elements represent protein similarity (TBLASTX, cut-off score of 75). CDSs involved in similar function are given the same colour: blue, regulation; white, non-ribosomal peptide biosynthesis; green, transport; grey, hypothetical or unclassified. Putative functions of CDS are indicated.

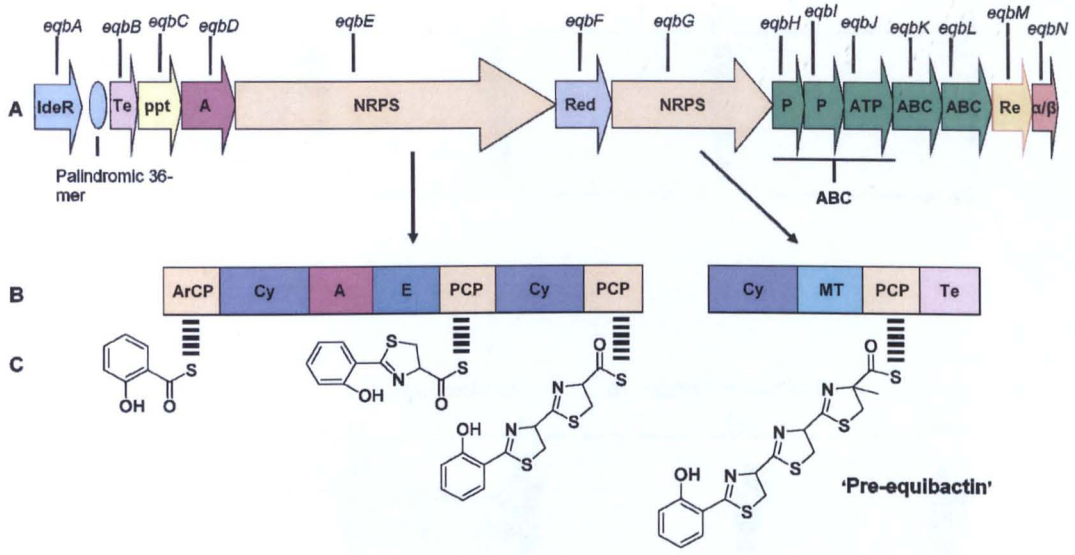


Figure 3.5 The equibactin locus and predicted functions of the *eqb* gene products

(A) Predicted functions of CDSs in the ICE*Se2 eqb* cluster (IdeR, iron dependent regulator; Te, type-II thioesterase; ppt, 4'-phosphopantetheinyl transferase; A, salicylate-AMP ligase; NRPS, non-ribosomal peptide synthetase; Red, thiazoline reductase; P, permease (component of ABC transporter); ATP, ATPase (component of ABC transporter); ABC, ABC transporter; Re, putative oxidoreductase; α/β, putative α/β hydrolase). See Table 3.2 for homology to other NRPS systems and transporters. (B) Organisation of modules and domains in the Eqb NRPS (ArCP, aryl acid carrier protein; Cy, heterocyclisation; A, adenylation; E, epimerization; PCP, peptidyl carrier protein; MT, methyl transferase; Te, type-I thioesterase. (C) Proposed intermediates in equibactin biosynthesis.

The prevalence of the equibactin locus in diverse *S. equi* and *S. zooepidemicus* populations was determined by Southern blot with a probe specific to the *eqbN* gene and PCR of the *eqbE* gene using primers listed in Table 3.1. The *eqbN* and *eqbE* genes were detected in 18 of 18 *S. equi* isolates but 0 of 73 *S. zooepidemicus* isolates. In addition, the location of the ICE*Se2* as determined by Southern blot (Figure 3.6) and PCR across flanking regions was identical in all 18 *S. equi* strains. Chapter 2 also describes the use of PCR to screen 12 more *S. equi* isolates (all positive) and another 102 *S. zooepidemicus* isolates (all negative) for the presence of the *eqbE* gene.

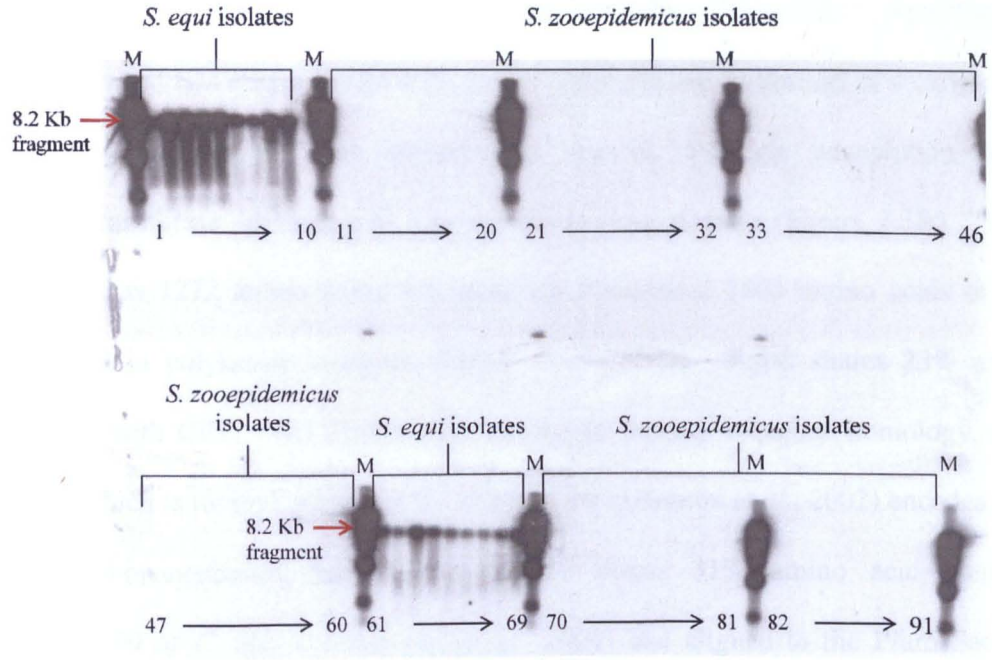


Figure 3.6 Southern blot showing the prevalence of the *eqbN* gene in diverse *S. equi* and *S. zooepidemicus* isolates

DNA was probed with a DIG-labelled 317 bp *eqbN*-specific DNA probe generated by PCR. M: DIG-labelled DNA molecular weight marker II (Roche Applied Science). See Appendix Table A.4 for details of strains tested.

3.4.2 Functions of the equibactin gene cluster as predicted by genetic analysis

Dissection of the modular structure revealed a remarkable similarity between the *C. kluyveri*, yersiniabactin and equibactin NRPS systems (Table 3.2). Yersiniabactin, a siderophore produced by *Yersinia* sp., is composed of one salicylate and 3 thiazoli(di)ne heterocycles derived from cysteine. A PKS module links the 2nd and 3rd heterocycles with a methylated malonyl-CoA derived linker (Miller *et al.*, 2002). The NRPS system of *C. kluyveri* is predicted to synthesise a similar aryl-capped siderophore containing salicylate and 4 thiazoli(di)ne rings derived from cysteine (Seedorf *et al.*, 2008). Siderophore activity has been demonstrated in the supernatant of *C. kluyveri*, but the biosynthetic cluster(s) responsible for siderophore production has not yet been established (Seedorf *et al.*, 2008). In *Se*4047, the first of the two peptide synthetases, EqbE, shares 45% and 30% aa sequence identity with CKL_1505 and HMWP2, respectively, while EqbG shares 25%

aa sequence identity with CKL_1511 and HMWP1 (Table 3.2). Analogous to the characterised HMWP1 and HMWP2, EqbG and EqbE are comprised of 4 carrier domains, 3 cyclisation domains, an epimerisation domain, a single adenylation domain, a methyltransferase domain and a type-I thioesterase domain (Figure 3.5B). The EqbG protein has 1272 amino acids, but lacks the N-terminal 1895 amino acids of HMWP1 involved in polyketide synthesis (Miller *et al.*, 2002). EqbC shares 23% amino acid identity with CKL_1523 (Table 3.2). Although lacking sequence homology, *eqbC* like *ybtD,* which is located outside of the *Y. pestis* HPI (Bobrov *et al.*, 2002) encodes a putative 4'-phosphopantetheinyl transferase. EqbM shares 31% amino acid identity with CKL_1509 of *C. kluyveri* (Seedorf *et al.*, 2008) and aligned to the Pfam Saccharopine dehydrogenase family, a member of the FAD/NAD(P)-binding Rossmann fold superfamily of redox enzymes. EqbN, a putative membrane protein has a predicted N-terminal transmembrane helix and shares 31% amino acid identity with a hypothetical protein Csac_2163 of *Caldicellulosiruptor saccharolyticus*. EqbN and Csac_2163 aligned to the Pfam α/β hydrolase family of enzymes that contain a catalytic triad associated with diverse catalytic activity (Table 3.2).

Table 3.2 BLASTP/FASTA analysis of the *eqb* cluster

S. equi CDS	CDS	Amino acid sequence identity %	Match by FASTA/BLASTP analysis Function	Organism	Accession no.	Reference
EqbA	gbs1749	42	Putative repressor	*S. agalactiae serotype III*	Q8E3K9	(Glaser *et al.*, 2002)
	CKL_1082	40	Putative repressor	*C.kluyveri*	A5N743	(Seedorf *et al.*, 2008)
	MntR	33	Manganese-dependant repressor	*B. subtilis*	P54512	(Que & Helmann, 2000)
	DtxR	28	Iron-dependant repressor	*C. diphtheriae*	P33120	(Qiu *et al.*, 1996)
EqbB	CKL_1503	52	Thioesterase	*C. kluyveri*	A5N8B4	(Seedorf *et al.*, 2008)
	PchC	34	Thioesterase (pyochelin biosythesis)	*P. aeruginosa*	P72176	(Serino *et al.*, 1997)
	ybtT	31	Thioesterase (yersiniabactin biosynthesis)	*Y. pestis*	Q56949	(Bearden *et al.*, 1997)
EqbC	GSP	27	4'-phospho-pantetheinyl-transferase	*Bacillus migulanus*	P40683	(Lambalot *et al.*, 1996)
	Sfp	25	4'-phospho-pantetheinyl-transferase	*B. subtilis*	P39135	(Quadri *et al.*, 1998)
	CKL_1523	23	Putative 4'-phospho-pantetheinyl-transferase (Sfp)	*C. kluyveri*	A5N8D4	(Seedorf *et al.*, 2008)
EqbD	CKL_1504	55	Putative salicyl-AMP ligase	*C. kluyveri*	A5N8B5	(Seedorf *et al.*, 2008)
	YbtE	42	Salicyl-AMP ligase (yersiniabactin biosynthesis)	*Y. pestis*	Q56950	(Gehring *et al.*, 1998a)
	DhbE	41	2,3-dihydroxy-benzoate-AMP ligase (bacillibactin biosynthesis)	*B. subtilis*	P40871	(May *et al.*, 2002)
EqbE	CKL_1505	45	NRPS (putative siderophore biosynthesis)	*C. kluyveri*	A5N8B6	(Seedorf *et al.*, 2008)
	HMWP2	30	NRPS (yersiniabactin biosynthesis)	*Y. pestis*	Q9Z399	(Gehring *et al.*, 1998b)
	Irp2	30	NRPS (siderophore biosynthesis)	*Photobacterium damselae* subspecies *piscicida*	Q2HQW9	(Osorio *et al.*, 2006)

Table 3.2 continued

			Match by FASTA/BLASTP analysis			
***S. equi* CDS**	**CDS**	**Amino acid sequence identity %**	**Function**	**Organism**	**Accession no.**	**Reference**
EqbF	YbtU	32 (partial)	Thiazoline reductase	*Y. pestis*	Q9Z3C6	(Gehring *et al.*, 1998a)
	CKL_1506	32 (partial)	Putative NRPS reductase	*C. kluyveri*	A5N8B7	(Seedorf *et al.*, 2008)
	CKL_1510	31 (partial)	Putative NRPS reductase	*C. kluyveri*	A5N8C1	(Seedorf *et al.*, 2008)
	Irp3	26 (partial)	Putative oxidoreductase (siderophore biosynthesis)	*P. piscicida*	Q2HQW7	(Osorio *et al.*, 2006)
	PchG	25 (partial)	NADPH-dependent thiazoline reductase (pyochelin biosynthesis)	*P. aeruginosa*	Q9HWG5	(Reimmann *et al.*, 2001)
EqbG	CKL_1511	25 (partial)	NRPS (putative siderophore biosynthesis)	*Clostridium kluyveri*	A5N8C2	(Seedorf *et al.*, 2008)
	HMWP1	25 (partial)	NRPS/PKS (yersiniabactin biosynthesis)	*Y. enterocolitica*	O54511	(Pelludat *et al.*, 1998)
	PchF	29 (partial)	NRPS (enantio-pyochelin biosynthesis)	*Pseudomonas fluorescens*	A8U045	(Youard *et al.*, 2007)
EqbH	Cbei_3098	54	Putative membrane protein	*Clostridium beijerinckii*	A6LY05	
EqbI	Cbei_3097	34	Putative cobalt transport membrane protein	*C. beijerinckii*	A6LY04	
	CKL_1516	24	Putative transporter (membrane protein)	*C. kluyveri*	A5N8C7	(Seedorf *et al.*, 2008)
EqbJ	Cbei_3096	48	Putative ABC transporter ATPase	*C. beijerinckii*	A6LY03	
	CKL_1517	45	Putative ABC transporter ATPase	*C. kluyveri*	A5N8C8	(Seedorf *et al.*, 2008)
EqbK	SAS2320	55	Putative ABC transporter (ATPase and permease)	*S. aureus*	Q6G6P2	(Holden *et al.*, 2004)
	Cbei_3094	48	Putative ABC transporter (ATPase and permease)	*C. beijerinckii*	A6LY01	
	CKL_1512	42	Putative ABC transporter (ATPase and permease)	*C. kluyveri*	A5N8C3	(Seedorf *et al.*, 2008)
	YbtP	29 (no match in N term 69 aa)	Yersiniabactin importer	*Y. pestis*	Q9ZG01	(Fetherston *et al.*, 1999)
	YbtQ	29 (no match in N term 109 aa)	Yersiniabactin importer	*Y. pestis*	Q9Z375	(Fetherston *et al.*, 1999)

Table 3.2 continued

S. equi CDS	CDS	Amino acid sequence identity %	Match by FASTA/BLASTP analysis Function	Organism	Accession no.	Reference
EqbL	SAS2319	57	Putative ABC transporter	*S. aureus*	Q6G6P3	(Holden *et al.*, 2004)
	Cbei_3093	46	Putative ABC transporter (ATPase and permease)	*C. beijerinckii*	A6LY00	
	CKL_1513	40	Putative ABC transporter (ATPase and permease)	*C. kluyveri*	A5N8C4	(Seedorf *et al.*, 2008)
	YbtP	28 (no match in N term 109 aa)	Yersiniabactin importer	*Y. pestis*	Q9ZG01	(Fetherston *et al.*, 1999)
	YbtQ	30 (no match in N term 125 aa)	Yersiniabactin importer	*Y. pestis*	Q9Z375	(Fetherston *et al.*, 1999)
EqbM	CKL_1509	31 (partial)	unknown	*C. kluyveri*	A5N8C0	(Seedorf *et al.*, 2008)
EqbN	Csac_2163	31 (partial)	unknown	*C. saccharolyticus*	A4XLG0	

The 3 predicted ABC transporters encoded by *eqbHIJ, eqbK* and *eqbL* are homologous (34-54% aa sequence identity) to the unknown ABC transporters in *Clostridium beijerinckii* (Cbei_3098, Cbei_3097, Cbei_3096, Cbei_3094 and Cbei_3093, respectively) (Table 3.2) and in *C. difficile* encoded on CDTn*2* (Figure 3.2, CD0429, CD0430, CD0431, CD0432, CD0433, 26-38 % amino acid sequence identity). In addition, EqbI, EqbJ, EqbK and EqbL share 24% to 45% amino acid identity with the putative *C. kluyveri* proteins CKL_1516, CKL_1517, CKL_1512 and CKL_1513, respectively (Seedorf *et al.*, 2008) (Table 3.2). Closer homologues to EqbK and EqbL include SAS2320 and SAS2319 in *S. aureus* (Holden *et al.*, 2004) (55 to 57% amino acid sequence identity) (Table 3.2).

Figure 3.7 Organisation of membrane spanning and ATP-binding domains in the *eqb* ABC transporters

Predicted TMHs are shown in white (a possible TMH is bordered by dashed lines); ATP-binding domains are shown in green.

Like ABC-type importers, the putative EqbHIJ transporter has transmembrane and ABC subunits as distinct polypeptide chains (Miethke & Marahiel, 2007) (Figures 3.5A, 3.7). *eqbH* and *eqbI* encode 2 permease domains containing 6 and 3 to 4 putative TMHs, respectively, whilst 2 ATP-binding components are encoded by *eqbJ* (Figure 3.7). However, no extracytoplasmic substrate binding protein CDS is apparent in the *eqb* cluster. Gram-positive bacteria tend to use a lipoprotein to scavenge the appropriate substrate and deliver it to the membrane channel of the associated ABC transporter. EqbI has a hydrophilic region located between the 3rd and 4th TMHs (Figure 3.7) that could potentially function as an extracellular substrate binding domain, although this region shows no homology to known ligand binding subunits. The cobalt uptake system of *S. typhimurium* represents a novel class of ATP-binding transporters that use extracytoplasmic regions within small integral membrane proteins (CbiN in *S. typhimurium*) to capture specific substrates (Rodionov *et al.*, 2006; Rodionov *et al.*, 2009). Other chimeric glycine betaine and glutamate/glutamine ABC transporters, which are widespread among Gram-positive bacteria, have fused substrate binding domains at the N or C terminus of translocator moieties (van der Heide & Poolman, 2002). Alternatively, the hydrophilic C-terminal region of EqbI may form a cytoplasmic loop and possible contact site for ATP-binding proteins (Davidson & Chen, 2004; Rodionov *et al.*, 2006). Most of the EqbI protein aligns poorly to the Pfam HMM of CbiQ and various other cobalt transport proteins (PF02361, score 0.00031). Although EqbI generally shares poor amino acid sequence identity with the Pfam members containing this domain, including CbiQ (12% identity), the domain organisation is similar. CbiQ is an unessential transmembrane component of the cobalt

uptake transport system in *S. typhimurium*, which also contains a possible cytoplasmic loop between the 4th and 5th predicted TMHs (Davidson & Chen, 2004; Rodionov *et al.*, 2006).

EqbK and EqbL, on the other hand, have a fused N-terminal permease and C-terminal ATPase, which would normally suggest involvement in efflux of extracellular toxins (Figure 3.7) (Chang & Roth, 2001; Chang, 2003; Dawson & Locher, 2006; Saurin *et al.*, 1999). Generally efflux ABC transporters possess 6 TMHs in each permease domain (Miethke & Marahiel, 2007); EqbK and EqbL contain 5 and 6 predicted TMHs, respectively (Figure 3.7). EqbK and EqbL may form the 2 chains of a single ABC transporter (heterodimer), or alternatively form 2 homodimeric ABC transporters. EqbK and EqbL are similar to the ABC transporters of *Y. pestis* (YbtP and YbtQ) and *M. tuberculosis* (IrtA and IrtB), which also have fused N-terminal permease and C-terminal ATPase domains typical of efflux systems. However, YbtP, YbtQ and IrtB function as ferric-siderophore importers (yersiniabactin and carboxymycobactin) rather than exporters and are therefore structurally unique among the subfamily of ABC transporters associated with iron transport (Farhana *et al.*, 2008; Fetherston *et al.*, 1999; Rodriguez & Smith, 2006). YbtP/YbtQ, and IrtA/IrtB were presumed to forms heterodimers (Fetherston *et al.*, 1999; Rodriguez & Smith, 2006) but IrtA and IrtB have recently been shown to form separate homodimers, with IrtA functioning as an exporter of apo-carboxymycobactin, not an importer of ferric-siderophore as previously thought (Farhana *et al.*, 2008). EqbK shares 29 to 27% amino acid sequence identity with YbtP and IrtA (not including the N-terminal 262 amino acids), respectively; EqbL shares 29 to 25% amino acid sequence identity with YbtQ and IrtB, respectively. EqbK and EqbL share 29% amino acid sequence identity with each other. These ABC transporters all show most homology in their C-terminal halves that encode the ATP-binding domain and contain the characteristic walker A, walker B and ABC transporter signature motifs. The N-terminal membrane components are more diverse, even when compared to each other (e.g. EqbK vs EqbL).

YbtP, YbtQ, IrtB, EqbK and EqbL are all similar in size (approximately 600 amino acids) but IrtA, which is larger, encodes an additional substrate binding domain in its N-terminal region (~260 amino acids). A siderophore interaction domain fused to the ABC transporter is an unusual feature. This particular domain is likely to be exposed on the intracellular surface and involved in scavenging the apo-carboxymycobactin to deliver it for export (Farhana *et al.*, 2008). No periplasmic/extracellular siderophore binding proteins are encoded in the *irt/ybt* clusters but IrtB is predicted to use Rv2985c, encoded elsewhere in the *M. tuberculosis* genome to bind ferric-carboxymycobactin (Farhana *et al.*, 2008; Fetherston *et al.*, 1999).

3.4.3 Substrate predictions

An important characteristic of the equibactin and yersiniabactin biosynthetic clusters is the presence of a single A-domain in HMWP2/EqbE that is probably responsible for the aminoacylation of all 3 different PCP domains (2 in HMWP2/EqbE and one in HMWP1/EqbG), 2 of which are not linked physically to the A-domain. By comparison to the GrsA A_{Phe} prototype, critical residues lining the substrate-binding pocket of the EqbE A-domain active site were identified (Challis *et al.*, 2000; Stachelhaus *et al.*, 1999) and found to match those mediating cysteine recognition with the exception of a change from asparagine to aspartate at residue 278 (Table 3.3). The *C. kluyveri* NRPS cluster has 2 putative cysteine-activating A-domains (Table 3.3) likely to aminoacylate 4 PCP domains, 2 of which are not linked physically to an A-domain.

Table 3.3 Prediction of EqbD and EqbE A-domain substrate specificity

		Residue (according to A_{Phe} numbering)								
A-domain[a]	Substrate	235	236	239*	278	299	301	322	330*	Accession number
HMWP2	Cys	D	L	Y	N	M	S	M	I	Q9Z399
BacA	Cys	D	L	Y	N	L	S	L	I	O68006
AngR	Cys	D	L	Y	N	M	S	M	I	P19828
PchE	Cys	D	L	F	N	L	S	L	I	Q9RFM8
PchF	Cys	D	L	Y	N	L	S	L	I	Q9RFM7
CtaC	Cys	D	L	Y	N	M	S	L	V	Q5MD35
MtaC	Cys	D	L	Y	N	M	S	L	I	Q9RFK9
BlmIV	Cys	D	L	Y	N	L	S	L	I	Q9FB18
CKL_1505	Cys?	D	L	Y	N	L	S	M	I	A5N8B6
CKL_1511	Cys?	D	L	Y	N	L	S	L	I	A5N8C2
EqbE	Cys?	**D**	**L**	**Y**	**D**	**M**	**S**	**M**	**I**	
DhbE	DHB	N	Y	S	A	Q	G	V	V	P40871
EntE	DHB	N	Y	S	A	Q	G	V	V	P10378
MxcE	DHB	N	F	S	A	Q	G	V	V	Q9F638
VibE	DHB	N	F	S	A	Q	G	V	V	O07899
AngE	DHB	N	F	S	A	Q	G	V	V	Q5DK17
YbtE	Sal	N	F	C	A	Q	G	V	L	Q56950
PchD	Sal	N	F	C	A	Q	G	V	I	Q9RFM9
CKL_1504	Sal?	K	F	C	A	Q	G	L	I	A5N8B5
EqbD	Sal?	**N**	**F**	**C**	**G**	**Q**	**G**	**I**	**I**	
SnbA	3-hydroxy-picolinic acid	N	F	C	S	Q	G	V	L	P95819

The predicted amino acid sequence between core A4 and A5 sequence motifs of the EqbE A-domain and EqbD were aligned, using CLUSTALW, to A-domains or aryl-amp ligase homologues with > 30 % sequence identity to EqbE or EqbD, respectively (Stachelhaus *et al.*, 1999). Based on the structural data of DhbE and GrsA, residues conferring substrate specificity were identified (May *et al.*, 2002; Stachelhaus *et al.*, 1999). * Amino acids at positions 239, 330 discriminate salicylate from DHB (May *et al.*, 2002). *S. equi* residues defined in bold differ from the consensus code of characterised substrate activating proteins. [a] Protein name; HMWP2, yersiniabactin-NRPS, *Y. pestis* (Gehring *et al.*, 1998b); BacA, bacitracin NRPS, *Bacillus licheniformis* (Konz *et al.*, 1997); AngR, anguibactin NRPS, *V. anguillarum* (Tolmasky *et al.*, 1993); PchE and PchF pyochelin NRPSs, *P. aeruginosa* (Quadri *et al.*, 1999); CtaC, cystothiazole A NRPS, *Cystobacter fuscus* (Feng *et al.*, 2005); MtaC, myxothiazol NRPS, *Stigmatella aurantiaca* (Silakowski *et al.*, 1999); BlmIV, bleomycin NRPS, *Streptomyces verticillus* (Du *et al.*, 2000); CKL_1505 and CKL_1511, *C. kluyveri* (Seedorf *et al.*, 2008); DhbE, bacillibactin DHB-AMP-ligase, *B. subtilis* (May *et al.*, 2002); EntE, enterochelin synthase, *E. coli* (Gehring *et al.*, 1997); MxcE, myxochelin DHB-AMP-ligase, *S. aurantiaca* (Silakowski *et al.*, 2000); VibE, vibriobactin DHB-AMP-ligase, *V. cholerae* (Wyckoff *et al.*, 1997); AngE, anguibactin DHB-AMP-ligase, *V. anguillarum* (Alice *et al.*, 2005); YbtE, yersiniabactin salicyl-AMP-ligase, *Y. pestis* (Gehring *et al.*, 1998b); PchD, pyochelin salicyl-AMP-ligase, *P. aeruginosa* (Quadri *et al.*, 1999); CKL_1504, *C. kluyveri* (Seedorf *et al.*, 2008); SnbA, pristinamycin I 3-hydroxy-picolinic acid-AMP-ligase, *Streptomyces pristinaespiralis* (de Crecy-Lagard *et al.*, 1997).

EqbD shares 42% amino acid identity to the YbtE salicylate-AMP ligase (Gehring *et al.*, 1998a), 41% sequence identity with DHB AMP ligase, DhbE (May *et al.*, 2002) and 55% amino acid identity with the uncharacterised CKL_1504 of *C. kluyveri* (Seedorf *et al.*, 2008). The adapted specificity conferring code for aryl acid activating domains discriminates between DHB and salicylate activating enzymes (May *et al.*, 2002). EqbD and CKL_1504 are predicted to activate salicylate since they both have a cysteine at position 239, which would impede access of the 3'-OH group of DHB and a more sterically demanding isoleucine at position 330 that replaces the conserved valine in DHB-activating enzymes (Table 3.3). However, a small change in EqbD from alanine to glycine was identified relative to residue 278 according to GrsA A_{Phe}-numbering, which differs from other NRPS systems that utilise salicylate. CKL_1504 has a lysine residue in place of the usual asparagine at position 235, which forms bivalent hydrogen bonds to the 2'- and 3'-OH groups of DHB in DhbE (May *et al.*, 2002).

3.4.4 Proposed mechanism of equibactin biosynthesis

3 condensation (C) domains identified in the 2 peptide synthetases, EqbE and EqbG are strongly modified, diverging from classical C domains involved in peptide bond formation. Instead, C-domains shared highest homology (29 to 32% amino acid sequence identity) with cyclisation (Cy) domains that catalyse thiazoline ring formation in yersiniabactin and bacitracin biosynthesis (Bobrov *et al.*, 2002; De Crecy-Lagard *et al.*, 1995; Konz *et al.*, 1997) or are predicted to catalyse similar heterocyclisation reactions in the uncharacterised NRPS system of *C. kluyveri* (24 to 57% amino acid sequence identity). 7 signature regions identified in Cy domains of *Y. enterocolitica* (yersiniabactin), *B. licheniformis* (bacitracin) and *V. anguillarum* (anguibactin) by (Konz *et al.*, 1997), were also mostly conserved in the 3 putative Cy domains of EqbE and EqbG (Figure 3.8) and the 4 putative Cy domains of the *C. kluyveri* NRPS system (data not shown). Some differences in these sequence

alignments were apparent although the contribution of residues in these regions to catalytic activity is not yet known. However, the highly conserved Cy motif DX_4DX_2S that corresponds in location to the highly conserved condensation domain catalytic core, HHX_3DGXS (Keating *et al.*, 2000; Konz *et al.*, 1997) is present in all 3 putative Cy domains of *S. equi* (Figure 3.8) and the 4 putative Cy domains of the *C. kluyveri* NRPS system (data not shown). This suggests an involvement of these putative Cy domains in heterocyclisation activity since the two asparate residues of this Cy core are essential to both amide bond formation and heterocyclisation by Cy1 of HMWP2 in *Y. pestis* (Keating *et al.*, 2000).

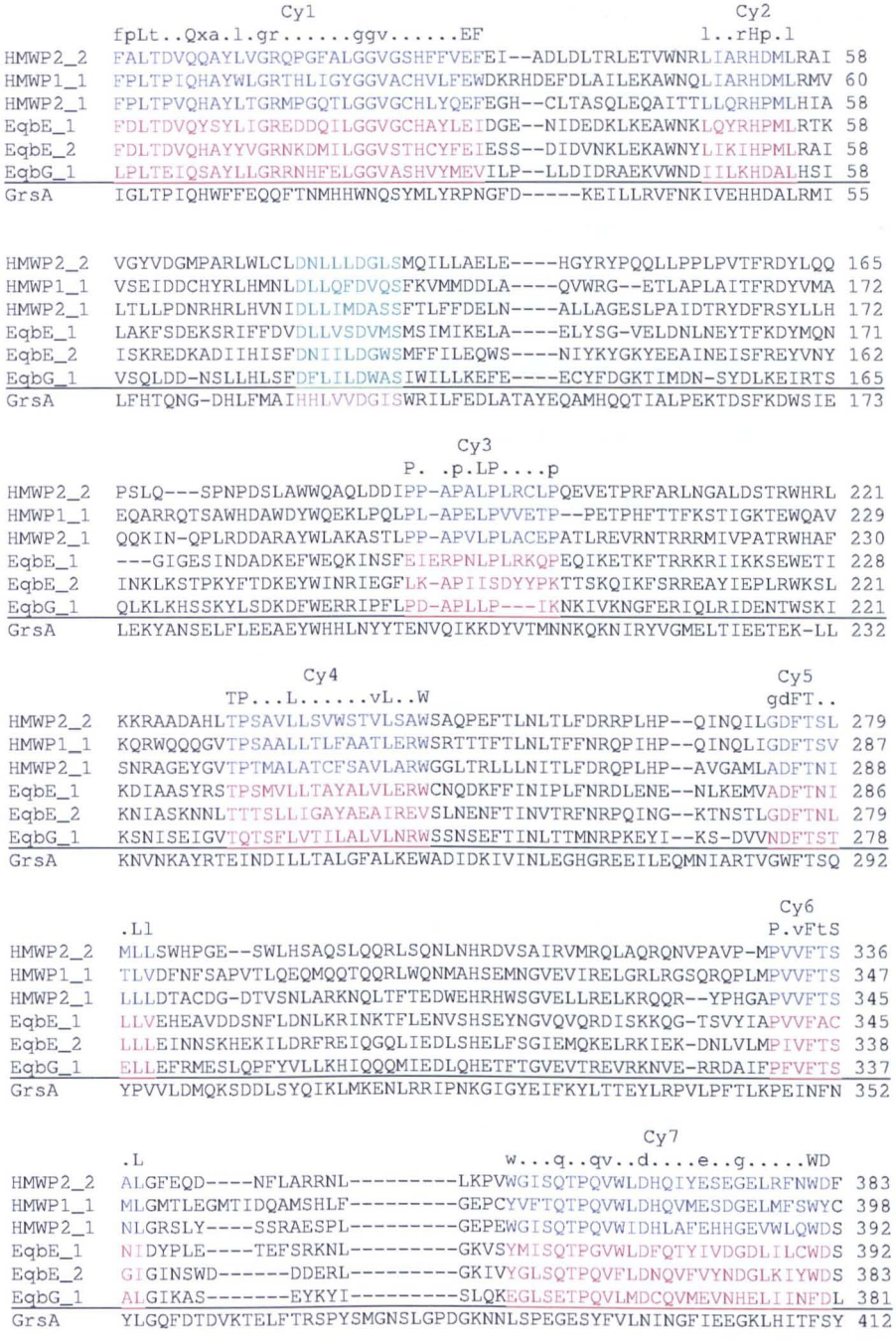


Figure 3.8 Bioinformatic prediction of *eqb* NRPS substrates

CLUSTALW alignment of amino acid sequences of the putative Cy domains of EqbE, EqbG *S. equi*, HMWP1, HMWP2 *Y. pestis* and the classical condensation domain of GrsA *B. brevis* (below line). Highlighted areas (red: *S. equi*, blue: *Y. pestis*) represent the 7 conserved signature sequences (C1-C7) with the consensus shown above (Konz *et al.*, 1997). The DX_4DX_2S catalytic core containing the 2 aspartate residues critical to amide bond formation and heterocylization (Keating *et al.*, 2000) is shown in green. The HHX_3DGXS catalytic core of classic condensation domains is shown in purple for GrsA (Konz *et al.*, 1997).

The presence of a putative thiazoline reductase (EqbF) with 32% amino acid identity with YbtU (Gehring *et al.*, 1998a) and 32% and 31% amino acid identity with CKL_1506 and CKL_1510 of *C. kluyveri*, respectively (Seedorf *et al.*, 2008), supports a prediction of thiazoline rings in the chemical structure of equibactin (Table 3.2). Thiazoline rings of yersiniabactin and pyochelin are reduced to thiazolidine by the activity of these thiazoline reductase enzymes (Miller *et al.*, 2002; Patel & Walsh, 2001). These enzymes also have a methyl transferase module, which although apparently non-functional in yersiniabactin biosynthesis, introduce a methyl group to the N-atom of the thiazolidine residue in pyochelin (Miller *et al.*, 2002; Patel & Walsh, 2001). A prediction of the NRPS intermediates is presented in Figure 3.5C.

Non-ribosomal peptides are usually released from the last carrier domain in the assembly line through the activity of a terminal thioesterase (type-I) (Challis & Naismith, 2004). The C-terminal domain of EqbG has similarity with type-I thioesterases. However, sequence alignment of this domain to thioesterases in Pfam showed a S1092 to A mutation in the critical serine residue (GXSXG-motif) of the predicted serine, histidine, aspartate catalytic triad essential for function (Bobrov *et al.*, 2002; Reimmann *et al.*, 2004). The terminal thioesterase of CKL_1511 (*C. kluyveri*) has an unusual GYSFSG motif, but the critical serine residue is conserved. Alignment of the type-II thioesterase encoded by *eqbB* showed an intact catalytic triad.

3.4.5 Regulation of the *eqb* NRPS

Upstream of the *eqb* operon is a CDS predicted to encode a putative metal ion dependent repressor, EqbA. Close homologues of *eqbA* and *eqbD* gene products are present in the genome of *S. agalactiae* serotype III NEM316, although the latter gene appears to be a pseudogene in this GBS. EqbA is predicted to have the characteristic N-terminal DNA binding, central metal ion binding and dimerisation domains of the MntR and IdeR family of manganese and iron-dependent repressors to which it has 33% and 28% sequence identity, respectively (Que & Helmann, 2000; Schmitt & Holmes, 1991). DtxR, a member of the growing family of IdeR global iron-dependent regulators in Gram-positive bacteria is activated on binding of divalent iron. Subsequent binding of the homodimeric form to a 21 bp DNA duplex blocks the transcription of downstream genes (Pohl *et al.*, 1999a; Schmitt & Holmes, 1991). A unique 36 bp incomplete palindrome (underlined) was identified in the promoter region of *eqbB* and represents a potential operator sequence for EqbA:-

5'-

AACTATTATTGTTAGATGTATCTAACAATAATAGTTCTAGTAG**TATATT**AATAATCAGATGGAAGGTGTTT

TGATG-3' (-35, –10 promotor sites and the ATG translational start are highlighted in bold).

To determine the role of EqbA on the regulation of the *eqb* NRPS, I generated a series of allelic replacement mutants in *Se*4047. Deletion of *eqbE* had no effect on the growth rate of *S. equi* on THA (Figure 3.9) or in THB (Figure 3.10B). However, deletion of *eqbA* produced very small colonies (Figure 3.9) as a result of a much reduced growth rate (Figure 3.10A). I hypothesised that this phenotype resulted from iron toxicity resulting from over-production of the product(s) of the *eqb* cluster. Over-expression of *eqbE* was confirmed since deletion of *eqbA* resulted in a 13-fold increase in *eqbE* transcript levels (Figure 3.11A). The generation of an Δ*eqbA*, Δ*eqbE* double deletion strain (Δ*eqbAE*),

which had a large colony phenotype on THA, established that the slow growth of the Δ*eqbA* strain was as a direct consequence of the function of the *eqbE* gene product (Figure 3.9). Transformation of the Δ*eqbA* or Δ*eqbAE* strains with the pGhost9 plasmid containing a 2nd copy of *eqbA* under the control of the *eqbA* promoter or a 2nd copy of *eqbE* under the control of the *eqbB* promoter complemented the Δ*eqbA* and Δ*eqbE* and induced a large or small colony phenotype, respectively (Heather *et al.*, 2008). The lack of an effect on the growth phenotype of *S. equi* grown *in vitro* following deletion of *eqbE* (Figures 3.9, 3.10) may be due to continued import of cations through the activity of several alternative cation transport systems, which include an HtsABC heme-binding system (SEQ0445 to SEQ0448) (Nygaard *et al.*, 2006), a putative MtsABC Mn^{2+} and Fe^{3+} metal transport system (SEQ1658 to SEQ1660) with 80 to 91% amino acid sequence identity to that of *S. pyogenes* (Janulczyk *et al.*, 1999) and a putative FtsABCD Fe^{3+} ferrichrome transport system (SEQ1836 to SEQ1839) with 59 to 77% amino acid sequence identity to that of *S. pyogenes* (Hanks *et al.*, 2005).

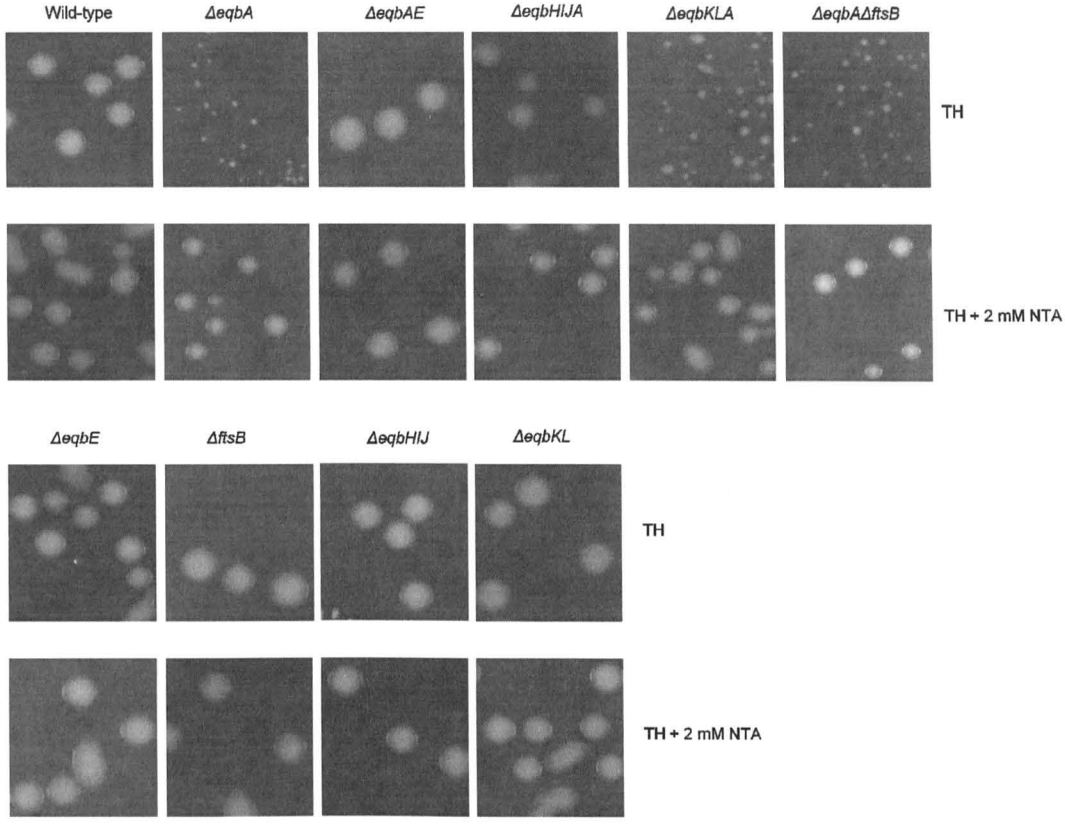
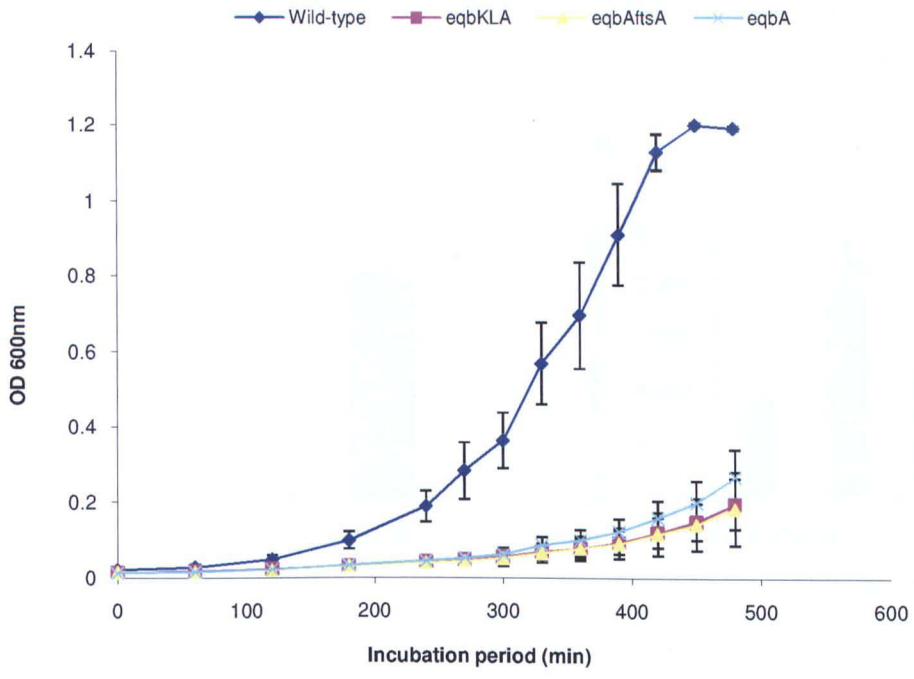


Figure 3.9 Influence of deletions in the *eqb* gene cluster on colony size

Photographs show colonies of wild-type, Δ*eqbA*, Δ*eqbAE*, Δ*eqbHIJA*, Δ*eqbKLA*, Δ*eqbAftsB*, Δ*eqbE*, Δ*ftsB*, Δ*eqbHIJ* and Δ*eqbKL*, *S. equi* strains grown overnight on THA and the increase in colony size of the Δ*eqbA*, Δ*eqbKLA* and Δ*eqbAftsB* strains grown on THA supplemented with 2 mM NTA.

A



B

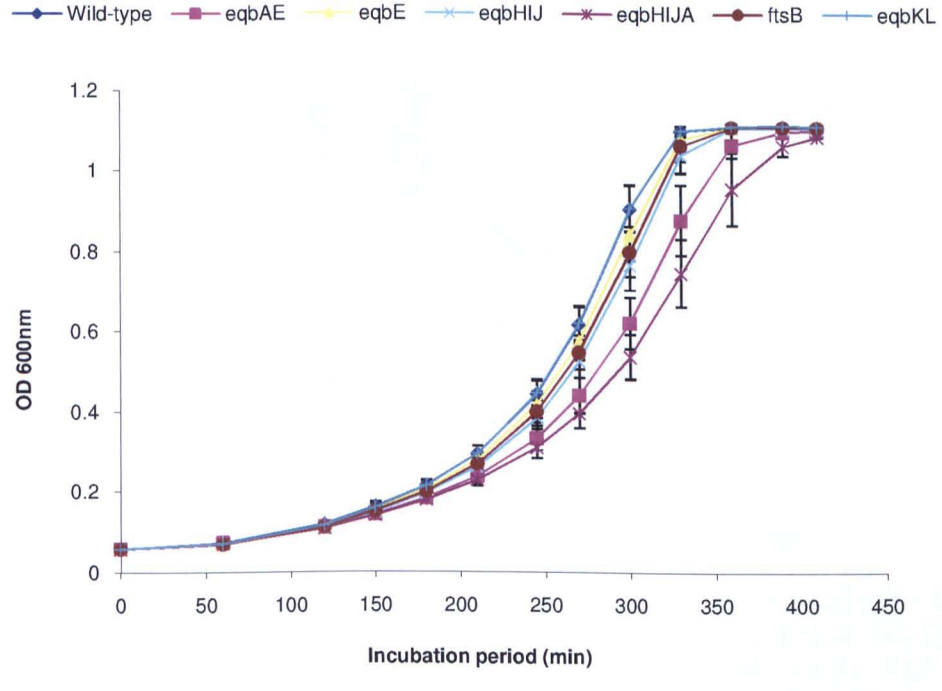


Figure 3.10 Effects of deletions in the *eqb* gene cluster on growth in THB

Graphs show the change in mean optical density at 600 nm over time (± SEM, n = 3). (A) Graph shows the slow growth of Δ*eqbKLA*, Δ*eqbA* and Δ*eqbAftsB S. equi* strains compared to the wild-type *S. equi* strain (using resuspended colony material as a starter culture). (B) Graph shows the similar growth rates of wild-type, Δ*eqbA*, Δ*eqbAE*, Δ*eqbHIJA*, Δ*eqbKLA*, Δ*eqbAftsB*, Δ*eqbE*, Δ*ftsB*, Δ*eqbHIJ* and Δ*eqbKL S. equi* strains (using overnight THB culture as a starter culture). The growth rate was slightly reduced in

the Δ*eqbAE* and Δ*eqbHIJA* strains compared to wild-type, with mean doubling times (± SEM) of 73.5 min ± 2.1, 79.5 ± 3.3 and 61.4 ± 0.7, respectively.

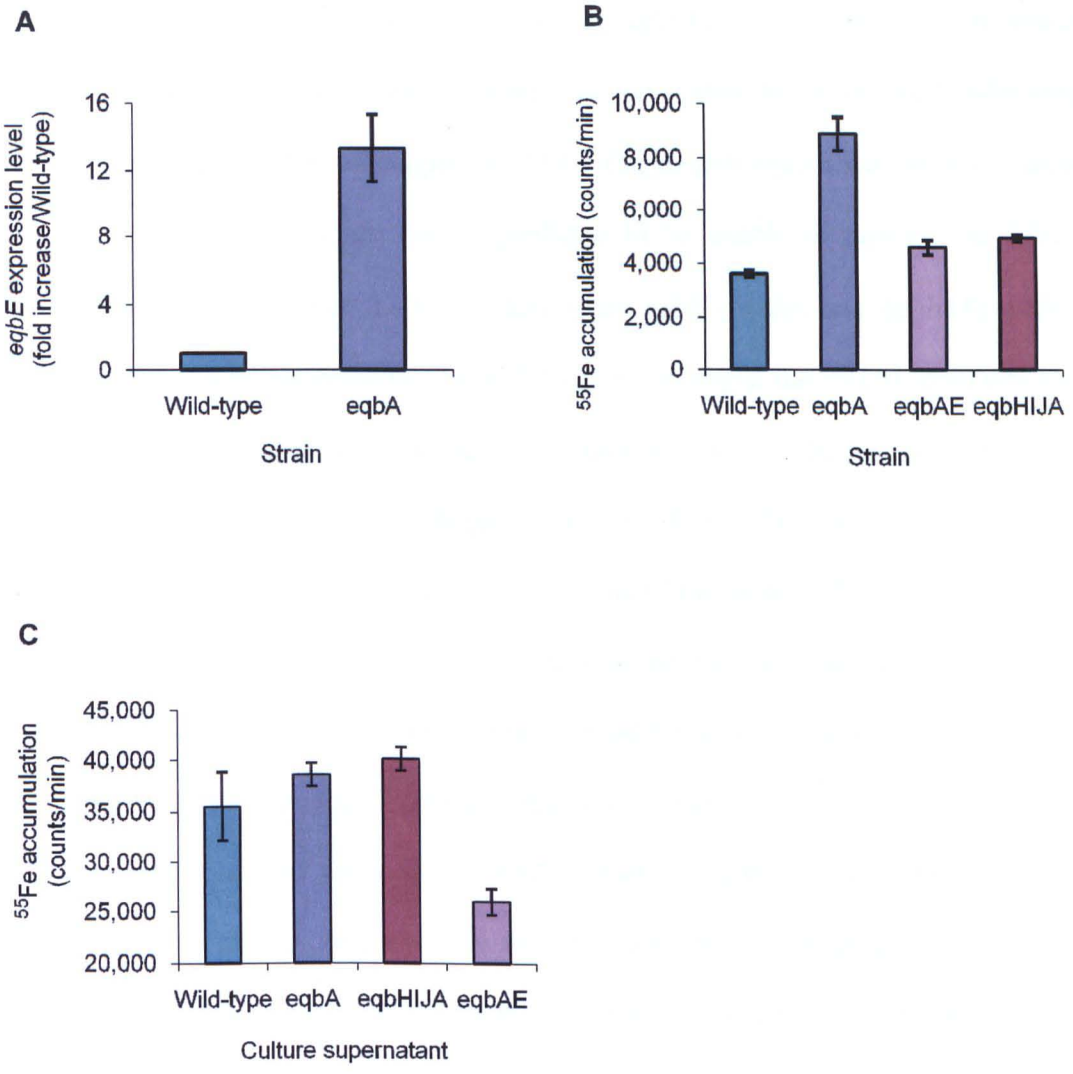


Figure 3.11 Phenotypic effects of deletions in the *eqb* gene cluster

(A) Fold increase in *eqbE* transcript level in the Δ*eqbA* strain relative to the wild-type *Se*4047 strain, which has been normalised to one (mean ± SEM, n = 3). (B) Quantification of ^{55}Fe accumulation by different *S. equi* strains (mean ± SEM, n = 3). The difference in ^{55}Fe accumulation between the Δ*eqbA* strain and the Δ*eqbAE,* Δ*eqbHIJA* and wild-type strains was found to be statistically significant using two-sample Wilcoxon rank-sum (Mann-Whitney) tests (P = 0.05, n = 3). (C) Quantification of ^{55}Fe accumulation by *S. equi* strain Δ*eqbAE* cross-fed with filter-sterilised culture supernatant from different *S. equi* strains grown to stationary phase.

Further studies supported the hypothesis that the *eqbA* deletion phenotype resulted from increased iron uptake. Supplementation of THA with 2 mM NTA, a known chelator of iron, restored near normal colony size in the Δ*eqbA* strain (Figure 3.9). Susceptibility to streptonigrin was used as an indirect measure of intracellular iron concentration (Brown *et al.*, 2002). Growth of wild-type *S. equi* was prevented by a minimal inhibitory concentration of 0.06 μM streptonigrin in THB. The Δ*eqbA* mutant was 16 times more sensitive to streptonigrin, whilst strains predicted to be unable to produce the NRPS product(s) through deletion of the biosynthetic gene *eqbE* (Δ*eqbE* and Δ*eqbAE*) were 2 times more resistant to the antibiotic (Table 3.4A). Interestingly, the *Sz*H70 strain was also twice as resistant to streptonigrin compared to wild-type *S. equi*, which is consistent with the lack of an *eqb* cluster in this subspecies (Table 3.4A). There was no difference in susceptibility of these strains to the antibiotic erythromycin. Therefore, increased sensitivity of the Δ*eqbA* mutant to streptonigrin compared to wild-type, Δ*eqbE* or Δ*eqbAE* strains suggests a greater intracellular iron pool in the former most likely as a result of de-repression of *eqb* non-ribosomal peptide synthesis. Similar differences in streptonigrin sensitivity were conferred when the Δ*eqbAE* mutant was cross-fed with filter-sterilised stationary-phase culture supernatant from the wild-type and various mutant strains (Table 3.4B), consistent with an accumulation of iron being mediated by an excreted, soluble factor.

Table 3.4 Production of the *eqb* NRPS product(s) by allelic replacement mutants of *Se*4047 and *Sz*H70

A

Strain	Streptonigrin MIC (μM)	Erythromycin MIC (μg ml^{-1})
Wild-type	0.06	0.016
Δ*eqbA*	0.004	0.016
Δ*eqbE*	0.125	0.016
Δ*eqbHIJ*	0.06	0.016
Δ*eqbKL*	0.06	0.016
Δ*ftsB*	0.06	0.016
Δ*eqbAE*	0.125	0.016
Δ*eqbHIJA*	0.03	0.016
Δ*eqbKLA*	0.004	0.016
Δ*eqbA*Δ*ftsB*	0.004	0.016
*Sz*H70	0.125	0.016

B

Conditioned THB from strain	Streptonigrin MIC of strain Δ*eqbAE* (μM)	Erythromycin MIC of strain Δ*eqbAE* (μg ml^{-1})
Wild-type	0.03	0.016
Δ*eqbA*	0.004	0.016
Δ*eqbE*	0.125	0.016
Δ*eqbAE*	0.125	0.016
Δ*eqbHIJ*	0.016	0.016
Δ*eqbKL*	0.06	0.016
Δ*ftsB*	0.03	0.016
Δ*eqbHIJA*	0.002	0.016
Δ*eqbKLA*	0.016	0.016
Δ*eqbA*Δ*ftsB*	0.004	0.016

C

Conditioned CDM from strain	Supplementation	Streptonigrin MIC of strain Δ*eqbAE* (μM)	Erythromycin MIC of strain Δ*eqbAE* (μg ml^{-1})
Δ*eqbAE*		0.125	0.03
Δ*eqbHIJA*		0.125	0.03
Δ*eqbAE*	10 μM salicylate	0.125	0.03
Δ*eqbHIJA*	10 μM salicylate	0.0005	0.03

(A) Sensitivity of wild-type and allelic replacement strains of *S. equi* to streptonigrin and erythromycin. MIC refers to the minimum inhibitory concentration of antibiotic required to prevent growth. The 2-fold difference in the streptonigrin MIC of wild-type versus Δ*eqbE* was found to be statistically significant using a 2-sample Wilcoxon rank-sum (Mann-Whitney) test (P = 0.008, n = 4). (B) Streptonigrin and erythromycin sensitivity in the Δ*eqbAE* strain cross-fed with filter-sterilised culture supernatant from wild-type and allelic replacement strains grown to stationary phase in THB. The 2-fold difference in Δ*eqbAE* streptonigrin MIC conferred by cross-feeding with Δ*eqbA* conditioned CDM relative to Δ*eqbHIJA* conditioned CDM was found to be statistically significant using a 2-sample Wilcoxon rank-sum (Mann-Whitney) test (P = 0.008, n = 4). (C) Streptonigrin and erythromycin sensitivity in the Δ*eqbAE* strain cross-fed with filter-sterilised culture supernatant from Δ*eqbAE* and Δ*eqbHIJA* allelic replacement strains grown to stationary phase in CDM ± 10 μM salicylate.

The increased production of the *eqb* NRPS product(s) in the Δ*eqbA* mutant was further supported by $^{55}FeCl_3$ incorporation assays, which showed an almost 2-fold increase in intracellular iron in the repressor deletion strain, compared to wild-type or the Δ*eqbAE* mutant (Figure 3.11B). Δ*eqbAE* cross-fed with filter-sterilised stationary-phase culture supernatant from the Δ*eqbA* strain but not the Δ*eqbAE* strain had a similar increase in $^{55}FeCl_3$ accumulation (Figure 3.11C).

A slight reduction in the growth rate of the Δ*eqbAE* mutant (mean doubling time 73.5 min ± 2.1 standard error mean (SEM)) compared to the Δ*eqbE* (63.2 ± 1.2) and wild-type (61.4 ± 0.7) *S. equi* strains (Figure 3.10B) suggests that the EqbA regulator has further impacts on growth independent of iron uptake, particularly since both Δ*eqbAE* and Δ*eqbE* mutants have similar intracellular iron pools as indicated by streptonigrin sensitivity.

As the global ferric uptake repressor Fur is an important regulator of siderophore utilisation and iron homeostasis in Gram-negative bacteria (Hantke, 2001), a Fur homologue (SEQ0255) was also deleted in *S. equi* to assess the influence of this protein on *eqb* non-ribosomal peptide biosynthesis. Deletion of the *fur* gene to generate a Δ*fur* mutant strain failed to alter the large colony wild-type phenotype suggesting that this Fur homologue in *S. equi* has no role in the repression of the *eqb* NRPS (data not shown).

3.4.6 The mechanisms for transport of the NRPS product(s)

The small colony phenotype of the Δ*eqbA* mutant was exploited to identify other CDSs likely to be involved in *eqb* product function. Deletion of the putative ABC transporters encoded by *eqbK* and *eqbL* or the *ftsB* gene, which lies outside the *eqb* locus and encodes a putative ferric-siderophore receptor (Clancy *et al.*, 2006; Hanks *et al.*, 2005), did not prevent the generation of the small colony phenotype on subsequent deletion of *eqbA*

(Figure 3.9). However, deletion of the *eqbH*, *eqbI* and *eqbJ* genes, followed by deletion of *eqbA* (Δ*eqbHIJA*) did prevent the generation of the small colony phenotype (Figure 3.9). Reduced levels of ^{55}Fe in the Δ*eqbHIJA* strain also suggested that the *eqbH*, *eqbI* and/or *eqbJ* gene products are essential for the majority of *eqb* product dependent iron accumulation (Figure 3.11B).

To determine if *eqbH*, *eqbI* and *eqbJ* were important to the export or import of the NRPS product(s), filter sterilised culture supernatant from mutant strains was added to the Δ*eqbAE* strain and its susceptibility to streptonigrin was quantified. Media from the Δ*eqbHIJA* strain increased the sensitivity of the Δ*eqbAE* strain by 64-fold compared with a 32-fold increased sensitivity conferred by media from the Δ*eqbA* strain (P = 0.008, Table 3.4B). Similarly, media from the Δ*eqbHIJ* mutant conferred 2-fold greater streptonigrin sensitivity to the Δ*eqbAE* strain compared with media cross-fed from the wild-type strain. These data suggest that the Δ*eqbHIJA* and Δ*eqbHIJ* strains secrete, but are unable to import the NRPS product(s) resulting in a build-up of this product in the culture media. In addition, Δ*eqbAE* cross-fed with filter-sterilised culture supernatant from the Δ*eqbHIJA* strain had an increase in $^{55}FeCl_3$ accumulation slightly greater, although not significantly so, than that observed when Δ*eqbAE* was cross-fed with supernatant from the Δ*eqbA* strain (Figure 3.11C). However, the Δ*eqbHIJA* strain grows slower than wild-type *S. equi* (Figures 3.9, 3.10B, mean doubling time 79.5 ± 3.3 vs 61.4 ± 0.7, respectively) and is slightly more sensitive to streptonigrin (2-fold compared to wild-type, Table 3.4A), which infers that this mutant still suffers some effects of iron toxicity and that other mechanisms may also contribute to import of the *eqb* product(s).

The Δ*eqbA* and Δ*eqbA*Δ*ftsB* mutants were equally sensitive to streptonigrin and conditioned media from these strains conferred the same streptonigrin sensitivity to the

cross-fed Δ*eqbAE* strain (Tables 3.4A and 3.4B). In contrast, Δ*eqbAE* cross-fed with supernatant from the Δ*eqbKLA* strain had a 4-fold lower sensitivity to streptonigrin suggesting that Δ*eqbKLA* is partially defective in the secretion of the NRPS product(s) (Table 3.4B). Conditioned media from the Δ*eqbKL* strain, but not the Δ*ftsB* mutant also conferred a 2-fold lower sensitivity to streptonigrin in the cross-fed Δ*eqbAE* strain relative to the Δ*eqbAE* strain cross-fed with wild-type supernatant (Table 3.4B). These data are consistent with a reduced secretion of the NRPS product(s) by the Δ*eqbKL* and Δ*eqbKLA* strains and suggest that one or both of the EqbKL ABC transporters contribute to the export of the non-ribosomal peptide(s). The Δ*eqbKLA* strain does however exhibit the same sensitivity to streptonigrin as the Δ*eqbA* strain, suggesting that these strains have similar levels of intracellular iron, despite reduced secretion of the NRPS product(s) in the former strain. It is possible that intracellular accumulation of the NRPS product(s) occurs in the Δ*eqbKLA* mutant, which removes iron from iron-dependent regulators and upregulates alternative iron import mechanisms.

3.4.7 The biochemical requirements for the *eqb* NRPS

A homologue of the salicylate synthases encoded by *ybtS* in *Yersinia sp.* (Miller *et al.*, 2002) or *pchA* and *pchB* of *Pseudomonas sp.* (Serino *et al.*, 1995) were not identified in *Se*4047. *S. equi* cultured in Todd-Hewitt media prepared from bovine heart infusion and a tryptic digest of animal tissue required no supplementation for equibactin production. However, production of the *eqb* product by the Δ*eqbHIJA* strain in chemically defined media (CDM) required salicylate supplementation consistent with salicylate being a substrate for the equibactin NRPS (Table 3.4C). The CDM contains 0.2 mM L-cystine.

3.4.8 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays were used to analyse recombinant EqbA (rEqbA) binding to a 227 bp DNA fragment (A) containing the upstream region of the *eqb* operon (P_{eqb}) (-237 to -11 bp) (Figures 3.1, 3.12). Unbound DNA ran as 2 distinct bands on non-denaturing polyacrylamide gel, which may be an artifact of the PCR method employed to generate the DNA probe, resulting in resolution of an additional single-stranded probe band that runs above the double-stranded band of interest (Varhimo *et al.*, 2007). Unbound DNA was unaffected by the presence of different cations (125 μM) (Lanes 1, 3, 5, 7, 9, 11, 13, 19 and 25). Binding of rEqbA caused a shift in the mobility of the lower band (Lane 2). Pre-treatment of rEqbA with EDTA resulted in a partial shift in the mobility of the lower band (Lane 4), which could be enhanced in the presence of additional 125 μM Fe^{2+} or 125 μM Zn^{2+} (Lanes 6 and 12) and to a lesser degree by 125 μM Mn^{2+} (Lane 8), but not by 125 μM Fe^{3+} or 125 μM Cu^{2+} (Lanes 10 and 14). The presence of the -73 bp to -38 bp DNA palindrome was essential for rEqbA binding as no shift was observed using DNA fragments containing the -237 to -73 bp (B) (Lanes 15 and 16) or -237 to -135 bp (C) (Lanes 17 and 18) upstream regions of the *eqb* operon. Incubation of a control DNA fragment (Pierce LightShift^R Chemiluminescent EMSA Kit) with rEqbA did not result in a shift in mobility (Lanes 22-24) and a control crude extract (Pierce LightShift^R Chemiluminescent EMSA Kit) did not bind to P_{eqb} (Lanes 19-21), suggesting that rEqbA binds specifically to P_{eqb}. In addition, pre-incubation with 3.8 pmol unlabelled P_{eqb} (190-fold excess) reduced the amount of P_{eqb} lower band shifted by rEqbA (Lanes 25-27).

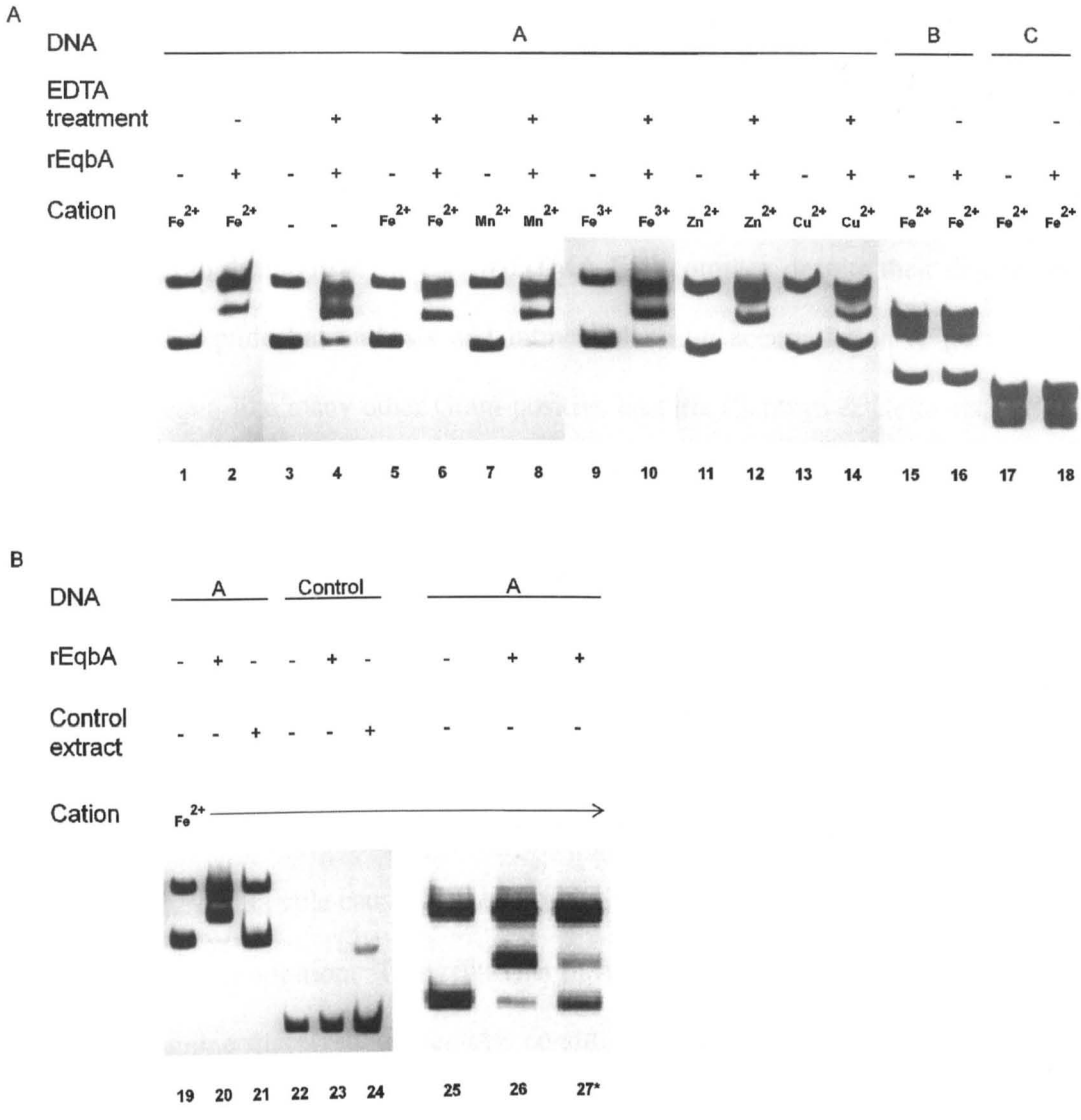


Figure 3.12 Electrophoretic mobility shift assay.

(A) Biotin end-labelled DNA fragments containing upstream regions of *eqbB*, A (227 bp P_{eqb}, -237 to -11 bp), B (165 bp, -237 to -73 bp) or C (103 bp, -237 to -135 bp) were incubated in the presence of 125 µM Fe^{2+}, 125 µM Mn^{2+}, 125 µM Fe^{3+}, 125 µM Zn^{2+}, 125 µM Cu^{2+} or control (-) as indicated with (+) or without (-) rEqbA and with (+) or without (-) prior treatment of rEqbA with EDTA. DNA fragments B and C lack the palindromic promoter region (-73 to -38 bp) immediately upstream of *eqbB*. Experimental conditions are described in the Methods. (B) DNA fragment A (P_{eqb}) or control DNA (Pierce LightShift^R Chemiluminescent EMSA Kit) were incubated in the presence of Fe^{2+} with (+) or without (-) rEqbA or control extract (Pierce LightShift^R Chemiluminescent EMSA Kit). * 3.8 pmol unlabelled DNA fragment A (P_{eqb}) was added 5 min prior to the addition of biotin-labelled P_{eqb} (lane 27).

3.4.9 Iron-chelating activity

The universal siderophore assay (Schwyn & Neilands, 1987) was used to compare the iron-chelating activity of various *S. equi* strains including Δ*eqbHIJA*, Δ*eqbA* and Δ*eqbAE*. Surprisingly, using a CAS agar overlay method, no differences were detected in the ability of these strains to decolourize the CAS-HDTMA-Fe^{3+} complex despite their disparities in non-ribosomal peptide biosynthesis and intracellular iron accumulation (Figure 3.13A). Growth of *S. equi,* like many other Gram-positive bacteria (Schwyn & Neilands, 1987), is inhibited by the detergent in the CAS-HDTMA-Fe^{3+} complex. As a result the *S. equi* strains were unable to grow in contact with the CAS agar. Growth of the control bacterium *S. aureus* was less effected by CAS-HDTMA-Fe^{3+} inhibition. Small but consistent differences were detected using the CAS supernatant assay on supernatant samples collected from Δ*eqbHIJA* and Δ*eqbAE* strains grown in CDM supplemented with salicylate (CDMs) (Figure 3.13B). Compared to supernatant from the Δ*eqbAE* strain, supernatant from the Δ*eqbHIJA* sample caused a reduction in absorbance (A_{630nm}) following incubation with the CAS assay solution. This reduction in A_{630nm} was equivalent to approximately 5 μM desferoxamine dissolved in Δ*eqbAE* conditioned media. However, when iron was omitted from the CDMs, these differences between Δ*eqbHIJA* and Δ*eqbAE* were no longer evident (even using the CAS shuttle assay) and no reduction in A_{630nm} occurred. Desferoxamine at a concentration of ≥ 2 μM dissolved in Δ*eqbAE* conditioned media (with iron omitted) did reduce A_{630nm} (data not shown). This finding suggests that the NRPS product(s) in Δ*eqbHIJA* conditioned medium was not able to remove iron from the CAS-iron complex directly, but may be able to bind free iron.

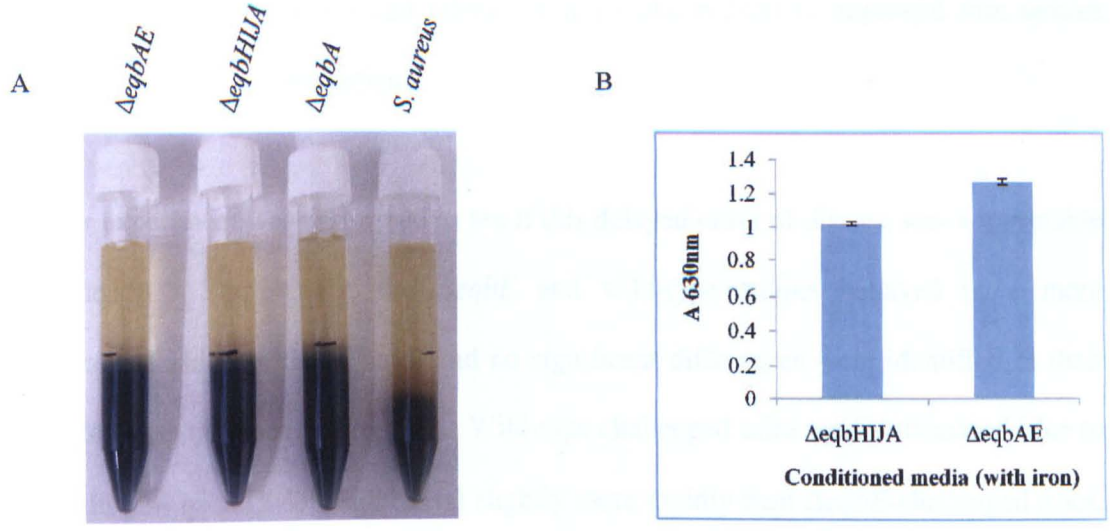


Figure 3.13 CAS siderophore assay

(A) Photograph shows an inability of *S. equi* mutant strains cultured as a suspension in THA (upper yellow layer) to remove iron from the CAS-iron complex (lower blue layer). *S. aureus* was included as a positive control. (B) Graph shows a reduction in absorbance at 630 nm in the CAS supernatant assay using conditioned media from the Δ*eqbHIJA* strain vs conditioned media from the Δ*eqbAE* strain. This difference between the Δ*eqbHIJA* and Δ*eqbAE* strains only occurred if iron was included in the chemically defined media supplemented with salicylate (CDMs).

3.4.10 Virulence of the Δ*eqbE* mutant strain in mice

In experiment one, the *eqbE* gene was shown to be essential for the full virulence of *S. equi* in mice (Figure 3.14). Δ*eqbE*-challenged mice had a significantly reduced weight loss (Figures 3.14A: $P < 0.0001$, B) and improved survival rate compared to wild-type-challenged mice (Figure 3.14F). 30% had gained weight 7 days post-challenge compared to just 5% of wild-type-challenged mice and 65% were still alive at day 7 post-challenge compared to only 45% of wild type-challenged mice. Significantly lower levels of *S. equi* were re-isolated from the nares of Δ*eqbE*-challenged mice ($P < 0.0001$) suggesting a reduced ability of this strain to colonise the respiratory tract (Figure 3.14E). However, although the onset of disease was delayed in the Δ*eqbE*-challenged mice, it was associated with deterioration in health as seen in the wild-type challenged group (Figures 3.14B, C

and D). Low levels of intracellular iron may have slowed down the early invasion of *S. equi* but could also have up-regulated other virulence factors able to aid directly or indirectly (via tissue damage and release of intracellular iron) in improved iron uptake through compensatory mechanisms.

Another experiment was performed to see if this delayed onset of disease was a repeatable phenomenon. Surprisingly, the Δ*eqbE* and wild-type strains behaved in a more comparable fashion in the 2nd trial and no significant differences were identified in their ability to cause disease (Figure 3.15). Wild-type challenged mice were euthanased (due to severe disease or > 15 % weight loss) slightly more rapidly than Δ*eqbE*-challenged mice, but by day 7 the number of surviving mice differed by just one between the 2 groups (Figure 3.15F). Pneumonia was detected in 4 extra mice in the Δ*eqbE*-challenged group compared to the wild-type-challenged group, but 2 of these 4 mice suffered from severe pneumonia compared to 4 in the wild-type-challenged group (Figure 3.15G). The occurrence of septicaemia was also similar for both groups (wild-type = 3, Δ*eqbE* = 2, Figure 3.15G). The prevalence of pneumonia and septicaemia has not been reported for experiment one because septicaemia was not measured in all cases and the accurate detection of pneumonia was complicated by euthanasia-induced changes in the lungs. However, pathological changes in the LNs were detected in one mouse from the Δ*eqbE*-challenged group in experiment one, which had a single LN abscess. In experiment 2, one wild-type-challenged mouse developed a LN abscess and 5 had enlarged LNs, whilst 3 mice had swollen LNs in the Δ*eqbE*-challenged group but no LN abscesses were detected. Overall, the pathological signs of disease were not significantly different between the 2 challenge groups.

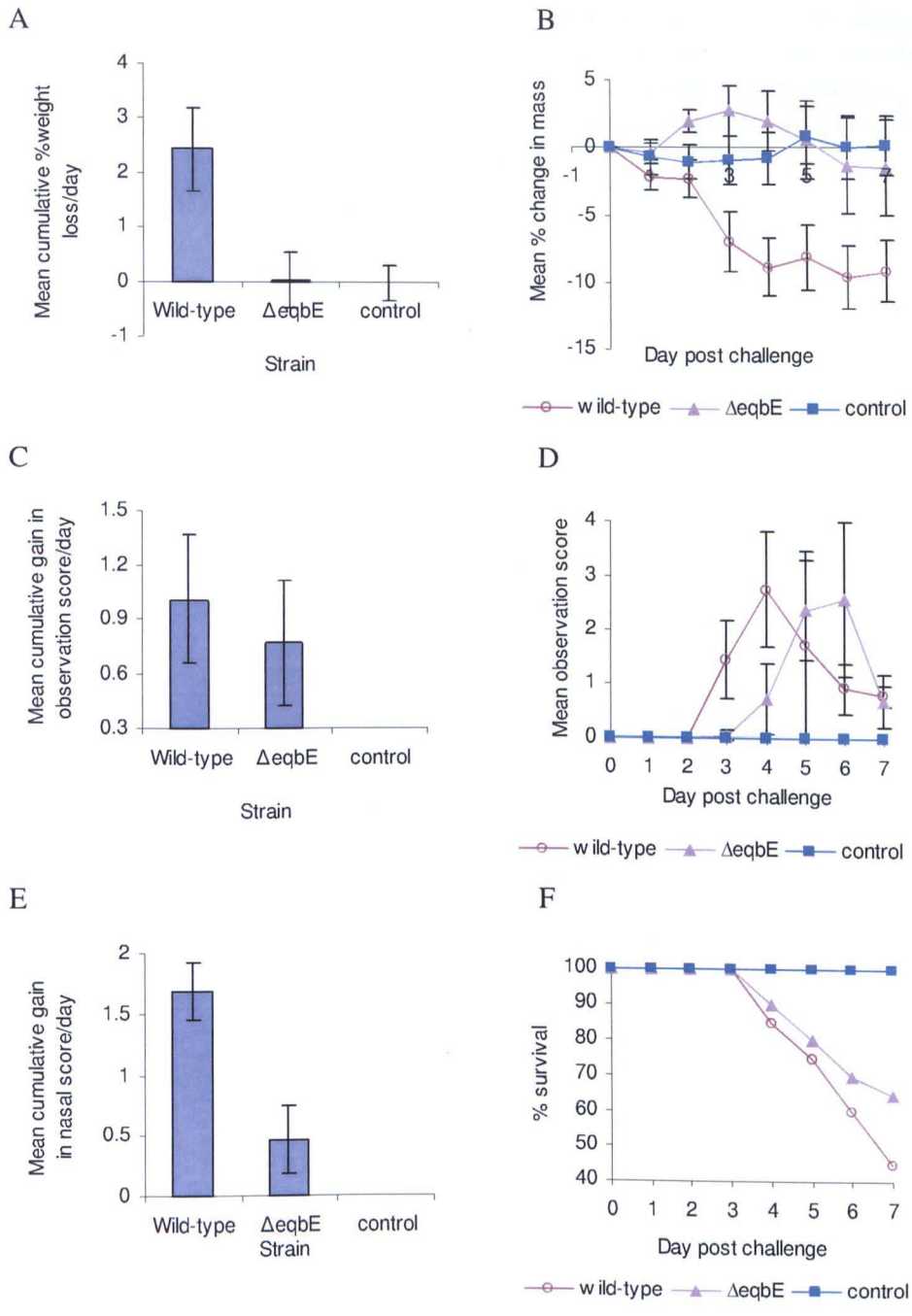
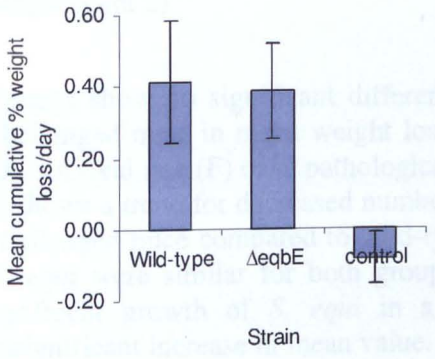


Figure 3.14 Virulence of the Δ*eqbE* mutant compared to wild-type *S. equi* in mice (experiment one)

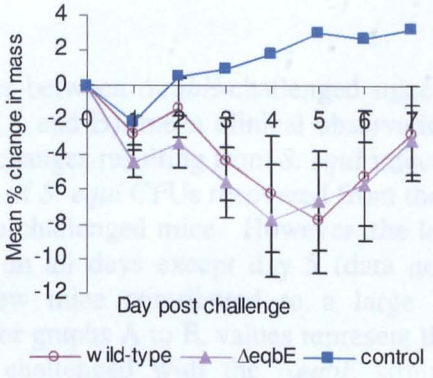
(A) Graph shows a significantly reduced mean cumulative percentage weight loss per day in the Δ*eqbE*-challenged mice compared to those challenged with wild-type *S. equi* ($P < 0.0001$). The measurement of cumulative weight loss per day takes into account the number of days survived in the post-challenge period and the influence of early euthanasia on mean values. (B) Graph shows a 2-3 day delay in the onset of weight loss in the Δ*eqbE*-challenged mice compared to wild-type-challenged mice. (C) Graph shows no significant difference in the mean cumulative gain in observation score between the 2 challenge groups (cumulative gain was calculated over the post-challenge survival period for each mouse). (D) Graph highlights the delayed increase in clinical observation score in the Δ*eqbE*-challenged mice compared to wild-type-challenged mice. (E) Graph shows that the mean cumulative nasal growth was significantly lower in the Δ*eqbE*-challenged mice

compared to wild-type-challenged mice ($P < 0.0001$). (F) Graph shows that Δ*eqbE*-challenged mice had a slightly better survival rate compared to wild-type-challenged mice. For graphs A to E, values represent the mean ± 95 % confidence interval. 20 mice were challenged with the Δ*eqbE* strain and 20 mice challenged with wild-type *S. equi*. 5 mice were present in the unchallenged control group.

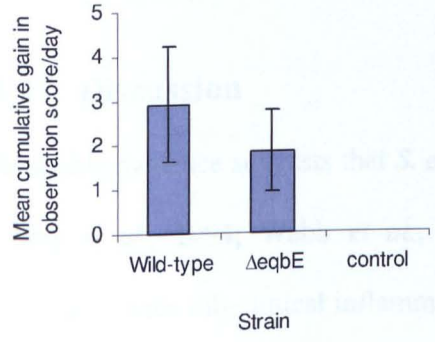
A


Mean cumulative % weight loss/day
0.60
0.40
0.20
0.00
-0.20
Wild-type
ΔeqbE
control
Strain

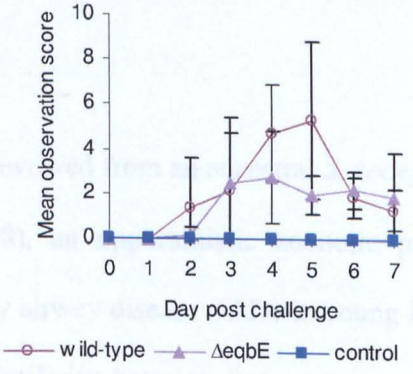
B


Mean % change in mass
4
2
0
-2
-4
-6
-8
-10
-12
0
1
2
3
4
5
6
7
Day post challenge
wild-type
ΔeqbE
control

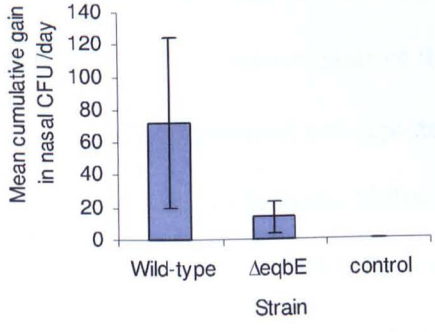
C


Mean cumulative gain in observation score/day
5
4
3
2
1
0
Wild-type
ΔeqbE
control
Strain

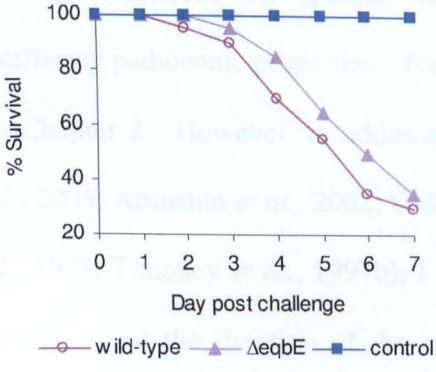
D


Mean observation score
10
8
6
4
2
0
0
1
2
3
4
5
6
7
Day post challenge
wild-type
ΔeqbE
control

E


Mean cumulative gain in nasal CFU /day
140
120
100
80
60
40
20
0
Wild-type
ΔeqbE
control
Strain

F


% Survival
100
80
60
40
20
0
1
2
3
4
5
6
7
Day post challenge
wild-type
ΔeqbE
control

G

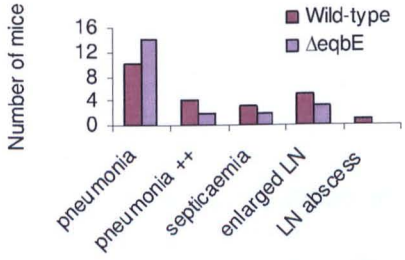

Number of mice
16
12
8
4
0
Wild-type
ΔeqbE
pneumonia
pneumonia ++
septicaemia
enlarged LN
LN abscess

Figure 3.15 Virulence of the Δ*eqbE* mutant compared to wild-type *S. equi* in mice (experiment 2)

Graphs show no significant differences between Δ*eqbE*-challenged mice and wild-type-challenged mice in mean weight loss (A and B), mean clinical observation score (C and D), survival rate (F) or in pathological changes resulting from *S. equi* infection (G). Graph E shows a trend for decreased numbers of *S. equi* CFUs recovered from the nares of *eqbE*-challenged mice compared to wild-type-challenged mice. However, the levels of isolated *S. equi* were similar for both groups on all days except day 5 (data not shown) when confluent growth of *S. equi* in a few mice contributed to a large but statistically insignificant increase in mean value. For graphs A to E, values represent the mean ± 95 % confidence interval. 20 mice were challenged with the Δ*eqbE* strain and 20 mice challenged with wild-type *S. equi*. 5 mice were present in the unchallenged control group.

3.5 Discussion

Available evidence suggests that *S. equi* evolved from an ancestral *S. zooepidemicus* strain (Jorm *et al.*, 1994; Webb *et al.*, 2008), an opportunistic zoonotic pathogen largely associated with sub-clinical inflammatory airway disease (IAD) in young horses (Wood *et al.*, 2005a). Given the close genetic similarity between these two streptococci, I used comparative genomics to identify differences between the species with a view to distinguishing key determinants of their differing pathogenic properties. The full results of this genome comparison are reported in Chapter 2. However, in addition to previously described genetic differences (Alber *et al.*, 2005; Artiushin *et al.*, 2002; Galan & Timoney, 1987; Lindmark *et al.*, 1999; Proft *et al.*, 2003; Timoney *et al.*, 1997b), I report here the identification of a streptococcal ICE, ICE*Se2* and the function of the siderophore-like NRPS system encoded by this element. This element was previously noted in a broad screen for ICEs in sequenced bacterial genomes, but no putative function was assigned (Burrus *et al.*, 2002a).

ICE*Se2* was present in all strains of *S. equi*, but absent from all strains of *S. zooepidemicus* examined. These strains were selected to be a diverse population based on their SeM allele (Kelly *et al.*, 2006; Waller & Jolley, 2007) or 16S-23S RNA gene intergenic spacer/SzP

PCR sub-type (Newton *et al.*, 2008) and were highly diverse based on their MLST sequence type (Webb *et al.*, 2008). My results also indicated that ICE*Se2* was located at the same genome position in all 18 *S. equi* strains, suggesting that its integration was an early and stable event in the evolution of *S. equi*.

The similarity of ICE*Se2* CDSs predicted to be involved in conjugation and site-specific recombination to those from *C. difficile* CDTn*2* and CDTn*5* is extremely interesting and raises the possibility of acquisition of the equibactin NRPS locus via a conjugative-like mechanism from a genetically unrelated bacterium. ICE*Se2* is also related to other ICEs in *Se*4047 (ICE*Se1*), *Sz*H70 (ICE*Sz1*) (Chapter 2) and *S. pyogenes* (ICE MGAS10750-RD.2) (Beres & Musser, 2007), athough these elements all carry distinct genetic cargo. These group A and group C streptococci may therefore share a common ICE reservoir. Although I predict integrative and conjugative genes are present in ICE*Se2*, to date I have failed to detect a circular intermediate by PCR through repeated experiments suggesting that this element may no longer retain the ability to recircularise and transfer to other bacteria. Similarly, circular forms of 4 out of 7 ICEs, including CDTn*2*, were not detected in *C. difficile* (Sebaihia *et al.*, 2006).

ICE*Se2*, CDTn*2* and CDTn*5* have putative conjugation modules that are closely related to the conjugation module of Tn*1549* in *Enterococcus spp.* (Garnier *et al.*, 2000) suggesting that they have inherited their similar structure from a common module ancestor. However, the serine recombinase encoded at the right flank of these elements differs from the Int (tyrosine recombinase) and Xis of Tn*1549*, which taken together with the differences in accessory functions highlights the importance of module exchange in the evolution of novel mosaic ICE structures. Serine recombinases are less common and genetically unrelated to the tyrosine recombinases and use a different mechanism for the site specific recombination of mobile elements (Grindley *et al.*, 2006). However, this family is rather

heterogeneous and ranges in size from 180 to nearly 800 amino acid residues. ICE*Se2* encodes a small recombinase (SEQ1229) with homology to the N-terminal half of the large TndX resolvase, which mediates excision and insertion of Tn*5397* by introducing 2 bp staggered cuts at the 3' ends of GA dinucleotides at the ends of the element (excision) or in the target site (integration) (Wang *et al.*, 2000). Consequently, direct GA dinucleotide repeats delineate the ends of Tn*5397* and one copy is also present at the joint of the circular form. CDTn*5* also has 5 bp direct repeats at these corresponding sites, whereas no direct repeats were identified in the flanks of ICE*Se2*. However, particularly large 108 bp perfect inverted repeats were present at the flanks of the element in *S. equi*, which like the much smaller imperfect inverted repeats (19-20 bp) in Tn*5397* and other ICEs may represent DNA binding sequences for the recombinase during the formation of the synaptic complex.

The sequence of a NRPS has been used previously to predict the structure of the peptide produced (Challis & Ravel, 2000; Lautru *et al.*, 2005). The similarity in type and organisation of functional modules in the equibactin NRPS and the yersiniabactin biosynthetic system (Miller *et al.*, 2002) leads me to suggest a tentative model for equibactin biosynthesis (Figure 3.5). Initially, I propose activation of the aryl acid carrier protein (ArCP) domain of EqbE and the peptide carrier domains (PCP) of EqbE and EqbG by the putative 4'-phosphopantetheinyl transferase encoded by *eqbC*. EqbD is proposed to activate salicylate, which is then transferred to the phosphopantotheine thiol of the ArCP-domain of EqbE. My observation that addition of salicylate is required for production of the *eqb* NRPS product(s) in chemically defined medium provides evidence in support of the proposed involvement of salicylate. Salicylate was also required for optimum production of the NRPS product(s) in *E. coli* containing the reconstituted NRPS (Chapter 4), but was not required for production of the NRPS product(s) by *S. equi* in Todd Hewitt media. This medium is prepared from bovine heart infusion and a tryptic digest of animal tissue and I propose that an appropriate NRPS substrate is released from host tissue during

infection with *S. equi*. The single adenylation (A)-domain of EqbE is predicted to activate cysteine, which is then transferred to the phosphopantotheine thiol of each of the 3 PCP-domains of the NRPS (2 in EqbE and one in EqbG). I predict that the cysteinyl thioesters attached to each PCP domain will be condensed and cyclised as the growing chain translocates to each of the 3 PCP domains. The presence of a putative thiazoline reductase (EqbF) suggests that as in yersiniabactin biosynthesis, the 2nd thiazoline ring of equibactin is likely to be reduced to a thiazolidine to prevent the auto-oxidation of the 1st ring to a thiazole. However, the available data do not rule out the possibility that the 1st or 3rd thiazoline ring could be reduced to a thiazolidine instead or as well.

Yersiniabactin is released from the last PCP-domain by a C-terminal thioesterase domain. However, the C-terminal thioesterase domain of EqbG has the critical serine residue (S1092) of the predicted serine, histidine, aspartate catalytic triad essential for function (GXSXG-motif) mutated to alanine. The type-II thioesterase encoded by *eqbB* has an intact catalytic triad, but to date these type-II TEs have only a presumed editing role in maintaining the efficiency of non-ribosomal peptide synthesis via removal of inappropriate substrates (Bobrov *et al.*, 2002; Butler *et al.*, 1999; Marahiel *et al.*, 1997; Reimmann *et al.*, 2004; Schneider & Marahiel, 1998). Interestingly, the tetrapeptide coelichelin is synthesised by a trimodular NRPS lacking a TE domain and is proposed to be released through the action of a separately encoded hydrolase (Lautru *et al.*, 2005). EqbN encodes a putative α/β hydrolase which could provide another potential mechanism for hydrolytic chain release similar to coelichelin. The *eqbN* gene was required for optimal production of the *eqb* NRPS product(s) by *E. coli* and induction of increased streptonigrin sensitivity in cross-feeding assays (Chapter 4). However, absence of *eqbN* did not abolish production of the NRPS product(s) in this system suggesting that *E. coli* may produce an endogenous hydrolase that can complement the activity of EqbN or that direct hydrolysis with water without the formation of an acyl enzyme intermediate is also possible, either at a low

background level or through a base-catalysed attack of water rather than nucleophilic attack of an active site serine (Li & Bugg, 2007).

Regulation of *eqb* non-ribosomal peptide synthesis is achieved through the action of the DtxR/IdeR-like transcriptional repressor EqbA. Deletion of *eqbA* led to increased transcription of the *eqbE* gene in the NRPS operon, over-production of the NRPS product(s) and increased import of iron causing toxicity and a small-colony phenotype. Deletions of *ideR* and other *ideR/fur* homologues have been lethal in *M. tuberculosis* and several species of *Pseudomonas*, *Vibrio* and *Neisseria* (Berish *et al.*, 1993; Tolmasky *et al.*, 1994; Venturi *et al.*, 1995). In *M. tuberculosis*, a rare deletion of *ideR* was obtained when the lethal effects of *ideR* inactivation were alleviated by a 2^{nd}-site suppressor mutation that restricted iron assimilation capacity (Rodriguez *et al.*, 2002). I propose that like other IdeR family members, EqbA is activated on binding iron, leading to homodimerisation and binding to a 36 bp DNA palindrome located immediately upstream of the NRPS operon to block its transcription. EqbA has a number of novel features compared with other IdeR family members. The suggested operator site is unlike other palindromic DNA iron boxes and this is reflected in the substitution of N-terminal Helix-Turn-Helix residues involved in DNA contact (Figure 3.16). Some of the conserved metal binding and dimerisation residues differ in EqbA. In addition, the SH_3-like fold of DtxR and IdeR, which functions to stabilize the DNA binding conformation of these repressors (Oram *et al.*, 2005; Pohl *et al.*, 1999b) is lacking in EqbA. EqbA is therefore likely to represent a new sub-class of the IdeR family. Electrophoretic mobility shift assays confirmed that EqbA does indeed bind to the *eqbB* promoter in a Fe^{2+}, Zn^{2+} and Mn^{2+} responsive manner. EqbA binding was not affected by Cu^{2+} or Fe^{3+} and was dependent on the presence of the *eqbB* promoter's DNA palindrome. The shift due to addition of Mn^{2+} was not complete and may reflect a reduced affinity of EqbA for Mn^{2+}. These data are in broad agreement with the activation of IdeR by different cations rather than MntR, which is Mn^{2+} specific (Chou *et al.*, 2004; Guedon

& Helmann, 2003; Que & Helmann, 2000; Schmitt *et al.*, 1995). Further comparison of the putative residues involved in the coordination of the regulatory metal (Guedon & Helmann, 2003) revealed that EqbA shares a cysteine at amino acid position 104 with DtxR, which confers Fe^{2+} responsiveness rather than the Mn^{2+} responsive glutamate possessed by MntR and TroR at the homologous position (Figure 3.16). The Fe^{2+} responsive methionine of DtxR has been replaced by asparagine at amino acid position 13 of EqbA, which is similar to the homologous residue in the Mn^{2+} responsive TroR.

```
EqbA (S.equi)  MNKIYHKEYVEQNYLENIYTLYLE--RDEVRNIDIVDKLGVARATVTHMLRNLEKKGYIK 58
DtxR (C.dip)   ---MKDLVDTTEMYLRTIYELEEE--GVTPLRARIAERLEQSGPTVSQTVARMERDGLVV 55
MntR (B.sub)   -----MTTPSMEDYIEQIYMLIEE--KGYARVSDIAEALAVHPSSVTKMVQKLDKDEYLI 53
TroR (T.pal)   --MSLVSDIAAENYLKTVVKALARSRRERVGTGELSRLLHVTPGTISTMVKRLEKGGYVQ 58
                          : *:. :     .          :   *     :::  : .:::   :

EqbA (S.equi)  YGDDKIVRFTSKGRTLAVELYEKHIYLTQVFK-HIGVDEKIAEIEACQIEHIISKDTFN- 116
DtxR (C.dip)   VASDRSLQMTPTGRTLATAVMRKHRLAERLLTDIIGLDINKVHDEACRWEHVMSDEVERR 115
MntR (B.sub)   YEKYRGLVLTSKGKKIGKRLVYRHELLDQFLR-IIGVDEEKIYNDVEGIEHHLSWNSIDR 112
TroR (T.pal)   RTHRLGCTLTRKGAVFGSAVLRKHRLLESFLSQVLCLEAGVVHKEAEMLEHACSDELIDV 118
                      :* .*  :.  :  :*     .:   : ::      :.   **  * :

EqbA (S.equi)  --KIKKYFE-----------------DKI------------------------------ 126
DtxR (C.dip)   LVKVLKDVSRSPFGNPIPGLDELGVGNSDAAVPGTRVIDAATSMPRKVRIVQINEIFQVE 175
MntR (B.sub)   IGDLVQYFE-----------------EDDARKKDLKSIQKKTEHHNQ------------- 142
TroR (T.pal)   IDRYLQYPTRDPHG-----------QPIPRKDTLLDLYVEDDVPGV------------- 153
                    :                   :

EqbA (S.equi)  ------------------------------------------------
DtxR (C.dip)   TDQFTQLLDADIRVGSEVEIVDRDGHITLSHNGKDVELIDDLAHTIRIEEL 226
MntR (B.sub)   ------------------------------------------------
TroR (T.pal)   ------------------------------------------------
```

Figure 3.16 CLUSTALW alignment of IdeRs

Primary structure alignment of EqbA homologues known to be regulated by Fe^{2+} or Mn^{2+}. Sequence alignment was generated using CLUSTALW. The putative residues involved in the coordination of the regulatory metal are shown in bold. The Helix-Turn-Helix motif is underlined. Sequences used are: DtxR, P33120 (Qiu *et al.*, 1996); MntR, P54512 (Que & Helmann, 2000); TroR, P96120 (Hardham *et al.*, 1997).

No iron-chelating activity could be detected in the Δ*eqbHIJA* or Δ*eqbA* mutants of *S. equi* using the universal CAS assay for iron chelators (Schwyn & Neilands, 1987) (Figure 3.13). Biochemists at Warwick University were also unable to identify equibactin by comparative metabolic profiling of the Δ*eqbA*, Δ*eqbAE* and Δ*eqbHIJA* strains using LC-MS. Furthermore, examination of biologically active conditioned LB (Luria-Bertani) and MM

(minimal media) from *E. coli* containing the reconstituted equibactin NRPS by LC-MS also failed to identify equibactin (Chapter 4). These data suggest that the concentration of equibactin in culture supernatant is low or the product of the *eqb* cluster may have a low affinity for iron and a structure different to that proposed. Pyochelin, a siderophore that lacks the 3rd incorporated thiazoline and malonyl linker of yersiniabactin, but otherwise shares the salicyl-bis-thiazonyl core has a relatively low affinity for iron. A role for pyochelin in the uptake of other essential metals has been proposed (Visca *et al.*, 1992) and this may also represent an alternative function for the *eqb* NRPS product(s). The host catecholamine norepinephrine, although unable to remove iron from CAS complex, can increase the availability of free iron to bacteria through an interaction that interferes with host glycoprotein iron sequestration (Freestone *et al.*, 2000).

These data are also consistent with a signaling role for the NRPS product(s) in the up-regulation of iron transport system(s). A number of siderophores are able to function as signaling molecules, influencing the expression of their own biosynthetic genes and cell surface receptors as well as other secreted virulence factors (Beare *et al.*, 2003; Crosa, 1997; Michel *et al.*, 2005; Michel *et al.*, 2007; Miethke & Marahiel, 2007; Pelludat *et al.*, 1998). Ferric-pyochelin acts as an intracellular effector via a direct interaction with an AraC-type regulator (Michel *et al.*, 2005). The AraC-type YbtA regulator also induces yersiniabactin operons involved in yersiniabactin biosynthesis, Fe-yersiniabactin uptake and salicylate synthesis, and the Fe-yersiniabactin receptor. An interaction of the YbtA with its DNA promoter target and Fe-yersiniabactin is proposed to be a prerequisite for the recruitment of the RNA polymerase complex (Anisimov *et al.*, 2005a; Anisimov *et al.*, 2005b; Fetherston *et al.*, 1996; Miethke & Marahiel, 2007; Perry *et al.*, 1999).

Siderophore-mediated iron uptake operons, similar to the ferric hydroxamate uptake (*fhu*) systems of *S. aureus* and *B. subtilis,* have been identified in other streptococci (Clancy *et*

al., 2006; Hanks *et al.*, 2005; Pramanik & Braun, 2006). FhuD, the associated lipoprotein receptor in *S. agalactiae* is able to bind a range of siderophores of both hydroxamate and catecholate classes. The homologous operon (42-55% amino acid sequence identity) is also present in the *S. equi* genome and I hypothesised that this could be involved in acquisition of the *eqb* NRPS product(s). However, deletion of the *ftsB* gene did not prevent iron toxicity due to overproduction of the NRPS product(s) on the further deletion of *eqbA* (Figure 3.9, Table 3.4). This indicates that FtsB is not absolutely required for import of the NRPS product(s) and further studies are necessary in order to determine if surface receptors play a role in import of the NRPS product(s).

EqbK and EqbL are ABC transporters with characteristics of ABC-type efflux systems (including siderophore export) but also structurally unique siderophore importers of *Y. pestis* and *M. tuberculosis* (Farhana *et al.*, 2008; Fetherston *et al.*, 1999). Although the Δ*eqbKLA* strain had a slow growth phenotype similar to the Δ*eqbA* strain (Figures 3.9, 3.10A) and both strains had the same sensitivity to streptonigrin suggesting equivalent intracellular iron stores (Table 3.4A), conditioned media from the Δ*eqbKLA* consistently conferred lower streptonigrin sensitivity to the cross-fed Δ*eqbAE* strain than media from Δ*eqbA* (Table 3.4B). Supernatant cross-feeding studies also demonstrated that strain Δ*eqbKL* conferred 2-fold decreased sensitivity to streptonigrin relative to wild-type (Table 3.4B). These conflicting results suggest that one or both of the *eqbKL* encoded ABC transporters plays a partial role in export of the NRPS product(s) and that the resultant accumulation of NRPS product(s) in the cell somehow impacts on iron and streptonigrin toxicity. A severe growth defect was seen in a strain of *B. bronchiseptica* that accumulated intracellular alcaligin due to a null mutation in a major facilitator superfamily transporter responsible for exporting this siderophore (Brickman & Armstrong, 2005). The drastically reduced growth rate was only seen when the strain was grown in conditions of high iron concentration and was reversed in a double deletion mutant that was also unable to

synthesise alcaligin (Brickman & Armstrong, 2005). The mechanistic basis for such an iron effect in the *B. bronchisep*tica mutant or the Δ*eqbKLA* strain of *S. equi* is unknown (Brickman & Armstrong, 2005). It is possible that intracellular accumulation of the *eqb* NRPS product(s) or alcaligin removes iron from iron-dependent regulators and upregulates alternative iron import mechanisms. Further work, including $^{55}FeCl_3$ accumulation assays will determine whether or not the Δ*eqbKLA* strain incorporates as much iron as the Δ*eqbA* strain. *eqbK* and *eqbL* may encode 2 polypeptide chains of a single ABC transporter heterodimer or as suggested for IrtA and IrtB of *M. tuberculosis*, may form separate ABC transporter homodimers with potentially distinct export/import functions (Farhana *et al.*, 2008). The generation of separate deletion mutants would help to clarify this. The ABC transporter encoded by the *eqbH*, *eqbI* and *eqbJ* genes plays a major role in the *eqb* dependent incorporation of iron.

Siderophore biosynthesis has not been identified in any streptococci examined to date (Eichenbaum *et al.*, 1996). A hybrid PKS-NRPS system has been noted in the published genome sequence of the oral pathogen *S. mutans*. Unlike the *eqb* locus, this NRPS module appears to incorporate 5 diverse residues into a molecule of unknown identity that is predicted to have more of an antibiotic-like structure (gramicidin/bacitracin family); expression from this locus was downregulated when the organism was grown in mixed species cultures (Huang, 2008). The only other *Streptococcus* strain displaying an NRPS is *S. thermophilus,* which encodes a single monomodular NRPS (Donadio *et al.*, 2007). However, homologues of *eqbA* and *eqbD* (pseudogene) are present in the genome of *S. agalactiae* serotype III, suggesting that a locus with similarity to the *eqb* NRPS may have been important to this organism at some point in its history.

Surprisingly, the mutant *S. equi* Δ*eqbE* strain that is unable to acquire iron though equibactin NRPS function was not significantly attenuated for virulence in mice, despite an

early experiment showing delayed onset of disease in this mutant strain compared to the wild-type strain (Figures 3.14, 3.15). The reasons for the disparities in virulence between the 2 mouse studies are unclear but highlight the need for reproducibility and caution when interpreting the results of such experiments, particularly when disease models in unnatural hosts are used. The mouse model used in this study resulted in a high incidence of pneumonia (Figure 3.15G), which is not a general feature of strangles in horses and may have limited the reliability of this method for establishing the virulence of the strain tested. Even studies looking at the virulence of mutant *S. equi* strains in mice using a less pneumonic-type disease model have given variable results when later compared with their ability to cause strangles in horses (Hamilton *et al.*, 2006). Quantification of the *in vivo* effects of the equibactin NRPS in horses may be the only way to establish its pathological significance.

3.6 Conclusions

ICE*Se2* is a key feature in the evolved genome of *S. equi*. Whilst I have not yet identified the peptide made by the equibactin NRPS or shown direct iron binding, data consistent with its role in iron acquisition has been established. Given the importance of iron acquisition to other streptococcal pathogens (Brown & Holden, 2002), the acquisition of ICE*Se2* may have contributed to the increased pathogenesis of this important streptococcus. Although the Δ*eqbE* mutant strain was not significantly attenuated in mice, the contribution of the equibactin NRPS to disease in the horse has yet to be established. More efficient acquisition of iron may enhance retropharyngeal LN abscessation and/or persistence in the GP of horses, both of which are critical to the establishment of long term carriage and vital to the success of *S. equi*.

In addition, this study has identified EqbA, a new member of the IdeR family of iron-dependent repressors. Transcriptional regulation of the equibactin NRPS is achieved by specific binding of EqbA to the *eqb* palindromic promoter sequence in an Fe^{2+}, Zn^{2+} and Mn^{2+} responsive manner. Further investigation is now warranted to gain a detailed understanding of the structure and mechanics of EqbA function and to establish the contribution (if any) of EqbA to the regulatory control of other potential virulence determinants.

Genes in the *eqb* cluster encode putative ABC transporters that contribute to the transport of the *eqb* NRPS product(s). EqbH, I and/or J have an essential role in *eqb* NRPS product(s) dependent iron import, whilst EqbK and/or EqbL may contribute to export of the secondary metabolite. Studies with further allelic replacement mutants containing deletions in individual and multiple transporter genes should shed more light on the precise function of each genetic component. Random mutagenesis may be a useful technique to establish any receptor requirements for the *eqb* ABC importer(s).

Salicylate is a proven substrate of the *eqb* NRPS. However, a concerted effort is clearly required to unravel the identity of the non-ribosomal peptide(s) synthesised by *S. equi*. The work detailed in Chapter 4 brings this important objective one step closer. Purification and elucidation of the *eqb* NRPS product structure is warranted to better understand the role played by the NRPS product(s) *in vivo*.

Chapter 4 Heterologous biosynthesis of equibactin utilising *Escherichia coli*

4.1 Introduction

NRPS systems produce peptides that have a composition of unusual amino acids and modifications, which often have tremendous therapeutic value. Harnessing this medicinal potential, however, is frequently limited by the rudimentary knowledge of growth requirements or poor growth characteristics of the original microbial host. In fact, most clinically important natural products, including many antibacterials, antifungals, immunosuppressants and antitumor agents, are produced by complex, slow growing microorganisms such as actinobacteria, myxobacteria and filamentous fungi (Challis, 2006). Consequently, there has been a move towards the heterologous production of these NRPS (and PKS or mixed NRPS-PKS) metabolites in rapidly growing, genetically tractable hosts like *E. coli*. Not only does the ease of *E. coli* fermentation makes this organism particularly suitable for metabolite overproduction but also, the genetic tractability of *E. coli* enables the use of genetic modification to manipulate and further study the biochemical process of secondary metabolite production. The modular nature of NRPS (and PKS) systems makes them amenable to the construction of recombinant hybrids and opens potential avenues for the manufacture of novel peptides (Donadio & Sosio, 2003; Watanabe *et al.*, 2003).

Examples of the *de novo* biosynthesis of NRPS or mixed PKS-NRPS products in *E. coli* include the production of diketopiperazine (an intermediate in the biosynthesis of the tyrocidine antibiotic produced by *B. brevis*) (Gruenewald *et al.*, 2004), the production of echinomycin (a quinoxaline antibiotic from *Streptomyces lasaliensis*) (Watanabe *et al.*, 2006) and the assembly of the thiazoline-containing siderophores pyochelin and

yersiniabactin from *P. aeruginosa* (Reimmann *et al.*, 2001) and *Y. pestis* (Pfeifer *et al.*, 2003), respectively.

Because of the large size or number of genes expressed, in most cases several plasmids containing mutually compatable origins of replication are required to assemble the gene clusters in *E. coli*. Most studies use natural or synthetic polycistronic operons under the control of an inducible T7 promoter to express the biosynthetic genes. The heterologous production of echinomycin, however, used 3 mutually compatible plasmids with 16 monocistronic gene cassettes, each with its own T7 promoter (Challis, 2006; Watanabe *et al.*, 2006); an approach that enabled the easy manipulation of gene clusters in order to examine the function of individual biosynthetic genes.

E. coli can also be utilized as a host for the overproduction and purification of recombinant biosynthetic proteins so that NRPS products can be synthesised in a test tube. This approach has been used to successfully reconstitute the 3 siderophores enterobactin (Gehring *et al.*, 1998c), pyochelin (Quadri *et al.*, 1999) and yersiniabactin (Miller *et al.*, 2002). *In vitro* reconstitution by this method coupled with HPLC and MS allows a thorough dissection of the molecular organization, mechanisms and substrates used by NRPS (and PKS) catalysts. For example, the *in vitro* production of a product matching the authentic yersiniabactin produced *in vivo* required the proteins: YbtE (salicylate adenylation domain), HMWP2 (NRPS), HMWP1 (PKS/NRPS), YbtU (reductase), Sfp (phosphopantetheinyl transferase) and substrates: salicylate, cysteine, malonyl-CoA, SAM, NADPH and ATP. Intermediate biosynthetic products were assessed by omitting catalytic domains from the reaction. For example, the function of the YbtU thiazoline reductase was deduced from MS analysis of reactions with and without YbtU. Deuterated substrates were used to validate the number of specific substrate molecules present in the final

product (Miller *et al.*, 2002). Such techniques can be used to build a picture of the complete biosynthetic process (Miller *et al.*, 2002).

4.2 Aims

S. equi is usually cultured in Todd Hewitt, a complex medium prepared from bovine heart infusion and a tryptic digest of animal tissue, supplemented with dextrose. A chemically defined medium (CDM) that supports the growth of *S. equi* was identified in this study (Chapter 3), but due to the metabolic requirements of streptococci, this CDM contained many components and remained complex in nature. Our biochemist collaborators at Warwick University were unable to identify a product from the ICE*Se2* NRPS and attributed this partly to the complex nature of the *S. equi* growth media. The work described in this chapter utilized *E. coli* for the production of the ICE*Se2* NRPS product(s) with the main aim of producing the NRPS product(s) in a less complex media in order to facilitate its identification by HPLC-MS analysis. By using different combinations of putative biosynthetic genes, I also intended to assess the contributions of individual proteins to the successful production of the secondary metabolite. The first objective was to clone all potential ICE*Se2* NRPS biosynthetic genes into compatible plasmids for their coordinated expression and the heterologous production of the NRPS product(s) in *E. coli* (cultured in LB broth or minimal media (MM)). A second objective aimed to reconstitute the ICE*Se2* NRPS product(s) *in vitro* using purified recombinant forms of the NRPS biosynthetic proteins.

4.3 Methods

4.3.1 Heterologous production of the ICE*Se2* NRPS product(s) in *Escherichia coli*

Novagen expression plasmids with compatible replicons were used for the co-expression of *eqbBCD*, *eqbMN*, *eqbF*, *eqbG* (multicystronic) and *eqbE* from T7 promotors in *E. coli* strain BL21(DE3) (Figure 4.1). DNA fragments containing *eqbBCD, eqbMN, eqbF, eqbG and eqbE* genes were amplified by PCR from *Se*4047 chromosomal DNA with Phusion polymerase (NEB), using primer pairs indicated in Table 4.1. *eqbBCD* was cloned into the multiple cloning site (MCS)-1 of pACYCDuet-1 using NcoI/BamHI to generate pACYC-BCD. *eqbMN* was cloned into MCS-2 of pACYC-BCD using BglII/AatII to generate pACYC-BCD-MN (chloramphenicol resistant). *eqbF* was cloned into MCS-1 of pCDFDuet-1 using NcoI/BamHI to generate pCDF-F. *eqbG* was cloned into MCS-2 of pCDF-F using FseI/AatII to generate pCDF-F-G (spectinomycin resistant). *eqbE* was cloned into pET21a using NheI/BamHI to generate pET21a-E (ampicillin resistant).

Figure 4.1 Plasmid contructs for the coordinate expression of 8 *eqb* genes in *E. coli*

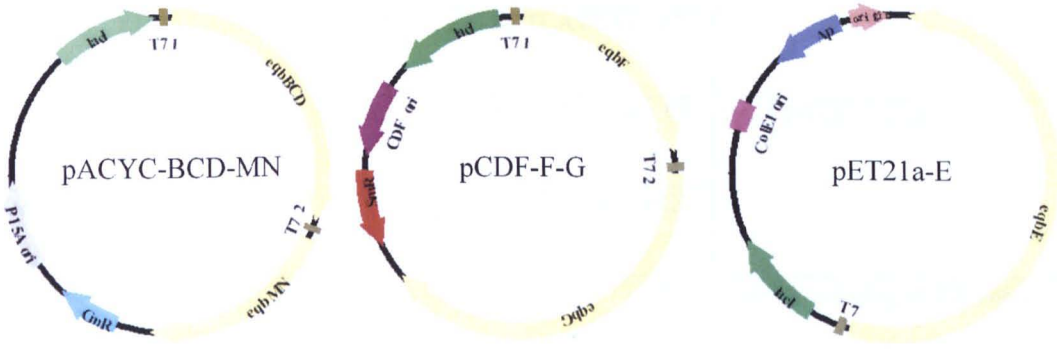


Table 4.1 Oligonucleotide primers used for cloning *eqb* gene expression vectors

Primer	Sequence (5'-3')	Purpose
ZM248	TAGTATCATCCATGGAAGGTGTTTTGATGGG	*eqbBCD* for cloning pACYC-BCD
ZM249	CACGACGGATCCAGTTCCATTCCAAAATCCT CC	
ZM402	GACGACGACCCATGGCAATGAAAATATATAG AATAGAAATTTC	*eqbC* fragment for cloning pACYC-CD-MN (from pACYC-BCD-MN)
ZM403	CGCAACTTCCGACTTATATGC	
ZM406	TGGTGTTAGTTACATCTAACAC	*eqbC* fragment for cloning pACYC-BC-MN (from pACYC-BCD-MN)
ZM407	CACGACGGATCCTCCTATATGCAAATTGCTC C	
ZM250	GACGACGACAGATCTTATGTATAATATAGGA GTTTTAGGTTG	*eqbMN* for cloning pACYC-BCD-MN and pACYC-MN
ZM251	GTAGGTAGCGACGTCCAGCCAATAAAACCCA CTAATG	
ZM311	GACGACGACAGATCTAATGGATATGAGTTAT TTGAATTTA	*eqbN* for cloning pACYC-BCD-N (used with primer ZM251)
ZM312	GTAGGTAGCGACGTCTTATAGATTAAGTGGT GATGTGC	*eqbM* for cloning pACYC-BCD-M (used with primer ZM250)
ZM252	GACGACGACCCATGGTTATGATGAAAAAAGT TATTATATGTGG	*eqbF* for cloning pCDF-F
ZM253	GACGGATCCCAAGACCTTTTCTGTACAACTC	
ZM254d	GACGACGACGGCCGGCCTCATGAAAATAGA GAATAAATTAATAGAG	*eqbG* for cloning pCDF-F-G and pCDF-G
ZM255	CACTACGCCGACGTCTTATTGCTCATTATTAT TCTCCTC	
ZM240	GACGACGACGCTAGCATGGAACTTAACAATA TAAAAGAAAG	*eqbE* for cloning pET21a-E
ZM241	GACGACGGATCCTTTTTTCATCATAAACTCCC TTC	
ZM313	GACGACGGATCCCGATGGAAGGTGTTTTGAT GGG	*eqbB* for cloning pGEX-*eqbB*
ZM314	GACGACCCCGGGCTTGTGTTAGATGTAACTA ACACC	
ZM315	GACGACGGATCCCAATGAAAATATATAGAAT AGAAATTTC	*eqbC* for cloning pGEX-*eqbC*
ZM316	GACGACGAATTCCTCCTATATGCAAATTGCT CC	
ZM317	GACGACGGATCCCTATGAATATAGAATTGAA AGAAAATTTAG	*eqbD* for cloning pGEX-*eqbD*

Table 4.1 continued

Primer	Sequence (5'-3')	Purpose
ZM318	GACGACGAATTCAGTTCCATTCCAAAATCCTCC	*eqbD* for cloning pGEX-*eqbD*
ZM319	GACGACGGATCCCAATGGAACTTAACAATATAAAAGAAAG	*eqbE* for cloning pGEX-*eqbE*
ZM320	GACGACCCCGGGCTTTTTTCATCATAAACTCCCTTC	
ZM321	GACGACGGATCCCTATGATGAAAAAAGTTATTATATGTGG	*eqbF* for cloning pGEX-*eqbF*
ZM322	GACGACGAATTCCAAGACCTTTTCTGTACAACTC	
ZM323	GACGACGGATCCCCATGAAAATAGAGAATAAATTAATAGAG	*eqbG* for cloning pGEX-*eqbG*
ZM324	GACGACCCCGGGTTATTGCTCATTATTATTCTCCTC	
ZM325	GACGACGGATCCCTATGTATAATATAGGAGTTTTAGGTTG	*eqbM* for cloning pGEX-*eqbM*
ZM326	GACGACGAATTCTTATAGATTAAGTGGTGATGTGC	
ZM376	GACGACGGATCCTTAAGATATTTAACAAAAGAATAGACTATG	*eqbN* for cloning pGEX-*eqbN*
ZM328	GACGACGAATTCCAGCCAATAAAACCCACTAATG	

Restriction sites are underlined.

pACYC-BCD-MN, pCDF-F-G and pET21a-E were introduced to BL21(DE3) via electroporation. The resultant *E. coli* strain was grown in Luria-Bertani (LB) broth media or a minimal media (MM) containing: 1.5 mg ml^{-1} KH_2PO_4, 4.34 mg ml^{-1} K_2HPO_4, 0.4 mg ml^{-1} $(NH_4)_2SO_4$, 0.22 mg ml^{-1} $MgSO_4.7H_2O$, 5 mg ml^{-1} glucose, 24.5 mg ml^{-1} $FeC_6H_5O_7$, 2.76 mg ml^{-1} $ZnSO_4.7H_2O$, 1 mg ml^{-1} $CaCl_2$, 2 mg ml^{-1} $Na_2MoO_4.2H_2O$, 1.21 mg ml^{-1} $CuSO_4$ and 0.5 mg ml^{-1} H_3BO_3. Media was supplemented with 50 μg ml^{-1} ampicillin, 50 μg ml^{-1} spectinomycin and 34 μg ml^{-1} chloramphenicol. Cultures were inoculated 4% (vol/vol) with a starter culture and growth was carried out at 37°C on a rotary shaker (220

rpm) to an optical density at 600 nm between 0.6 and 0.8. 0.4 mM IPTG was added together with 1 mM salicylate and the culture was incubated for a further 20 h at 28°C with shaking (220 rpm). Samples were taken for SDS-PAGE analysis prior to IPTG induction and 4 h post IPTG induction. After 20 h, supernatant was collected and filter sterilised for LC-MS analysis (see Chapter 3 methods). Supernatants were also tested for their ability to decolourise CAS reagent (Schwyn & Neilands, 1987) (see Chapter 3 methods). In addition, CAS test agar (Schwyn & Neilands, 1987) containing 1.2% agar, CAS indicator and MM, supplemented with salicylate, antibiotics and IPTG (as described above) was prepared to see if halos of decolourization appeared around cultured *E. coli* strains expressing all 8 Eqb proteins. 10 ml CAS indicator containing 5 ml CAS solution (0.06 g CAS in 50 ml H_2O), 1 ml iron (III) solution (1 mM $FeCl_3$ in 10mM HCl) and 4 ml HDTMA (0.07 g in 40 ml H_2O) was added per 100 ml of CAS test agar. For a positive control, *P. aeruginosa* was cultured on CAS test agar.

In order to test the influence of cross-fed filter sterilised supernatant from *E. coli* on the streptonigrin sensitivity of *S. equi* strain Δ*eqbAE* (see Chapter 3 methods), cells induced for 4 h with IPTG were harvested, washed 3 x and re-suspended in fresh media without antibiotics. Cells were incubated for another 16 h at 28°C with shaking (220 rpm) before the supernatant was collected, filter sterilised and analysed for its influence on streptonigrin sensitivity (Δ*eqbAE* cross-feeding bioassay) and by LC-MS.

Additional combinations of plasmids were constructed in order to assess the contribution of EqbB, EqbD, EqbE, EqbF, EqbG, EqbM and EqbN to the biosynthesis of the *eqb* NRPS product(s) in *E. coli* and also to provide appropriate negative controls for this system of production. An *eqbC* N-terminal DNA fragment was used to replace *eqbB* and part of *eqbC* in pACYC-BCD-MN following NcoI/NsiI digestion to generate pACYC-CD-MN (Table 4.1). An *eqbC* C-terminal DNA fragment was used to replace *eqbD* and part of

eqbC in pACYC-BCD-MN following NsiI/BamHI digestion to generate pACYC-BC-MN (Table 4.1). *eqbM* and *eqbN* were cloned separately into the MCS-2 of pACYC-BCD using BglII/AatII to generate pACYC-BCD-M and pACYC-BCD-N, respectively (Table 4.1). 2 *E. coli* strains lacking EqbB and EqbD contained pACYC-CD-MN, pCDF-F-G, pET21a-E and pACYC-BC-MN, pCDF-F-G, pET21a-E, respectively. 2 further *E. coli* strains lacking EqbE and EqbF contained pACYC-BCD-MN, pCDF-F-G, pET21a and pACYC-BCD-MN, pCDF-G, pET21a-E, respectively. An *E. coli* strain lacking EqbG contained pACYC-BCD-MN, pCDF-F and pET21a-E. Another 2 strains lacking EqbM or EqbN contained pACYC-BCD-N, pCDF-F-G, pET21a-E or pACYC-BCD-M, pCDF-F-G, pET21a-E, respectively.

4.3.2 *In vitro* reconstitution of the ICE*Se2* NRPS product(s)

The pGEX-3X expression system (GE Healthcare) was used for the production and purification of recombinant EqbB, EqbC, EqbD, EqbE, EqbF, EqbG, EqbM and EqbN proteins in *E. coli* DH10B using the method described in Chapter 3 (for the overexpression and purification of EqbA). Carl Robinson (AHT) expressed and purified the Eqb proteins. DNA fragments containing *eqbB, eqbC, eqbD, eqbE, eqbF, eqbG, eqbM and eqbN* genes were amplified by PCR from *Se*4047 chromosomal DNA with Phusion polymerase (NEB), using primer pairs indicated in Table 4.1. *eqbB, eqbE, eqbG* were cloned into the MCS of pGEX-3X using BamHI/SmaI digestion and *eqbC, eqbD, eqbF, eqbM, eqbN* were cloned into the MCS of pGEX-3X using BamHI/SmaI digestion. The resulting constructs were named pGEX-*eqbB* to pGEX-*eqbN* (Table 4.1).

The *in vitro* synthesis of the ICE*Se2* NRPS product was performed using the method used to reconstitute yersiniabactin (Miller *et al.*, 2002), with some modifications. A solution of 0.2 μM EqbB, 0.2 μM EqbC, 0.2 μM EqbD, 0.2 μM EqbE, 0.2 μM EqbF, 0.2 μM EqbG,

0.2 μM EqbM, 0.2 μM EqbN, 100 mM Tris-Cl (pH 7.6), 0.1 M NaCl, 0.8 mM $CaCl_2$, 1 mM TCEP, 1 mM CoASH, 2 mM L-cysteine, 0.75 mM SAM, 0.75 mM NADPH and 2mM salicylate was incubated at 30°C for 1 hr to allow phosphopantetheinylation of the EqbE and EqbG. As a control, the same reaction mixture was prepared, but with EqbE and EqbG omitted. Then, 10 mM ATP was added, and the reaction mixtures were incubated for 14 hr at 30°C. All reagents were purchased from Sigma. Reaction mixtures were stored at -20°C until their analysis by LC-MS according to the methods described in Chapter 3. A small sample was diluted 1 in 4 (in water), filter sterilized and analysed for its influence on streptonigrin sensitivity using the Δ*eqbAE* cross-feeding bioassay (see Chapter 3 methods). The diluted sample was also tested for its ability to decolourise CAS reagent (see Chapter 3 methods).

4.4 Results

4.4.1 Heterologous production of the ICE*Se2* NRPS product(s) in *Escherichia coli*

To determine if all of the genes necessary for the production of the ICE*Se2* NRPS product(s) were present in the *eqb* locus, and to produce this product in a less complex media, *eqbB*, *eqbC*, *eqbD*, *eqbE*, *eqbF*, *eqbG*, *eqbM* and *eqbN* were cloned into 3 expression plasmids with compatible replicons (Figure 4.1). Production of all 8 Eqb proteins was induced by IPTG treatment of transformed *E. coli* strain BL21(DE3) cells grown to log phase in LB broth or a MM in the presence of 1 mM salicylate as described in Methods. SDS-PAGE analysis revealed protein bands corresponding to EqbD, EqbF and EqbE in induced samples (Figure 4.2); however protein bands representing the other 5 Eqb proteins were not obvious in the soluble fraction.

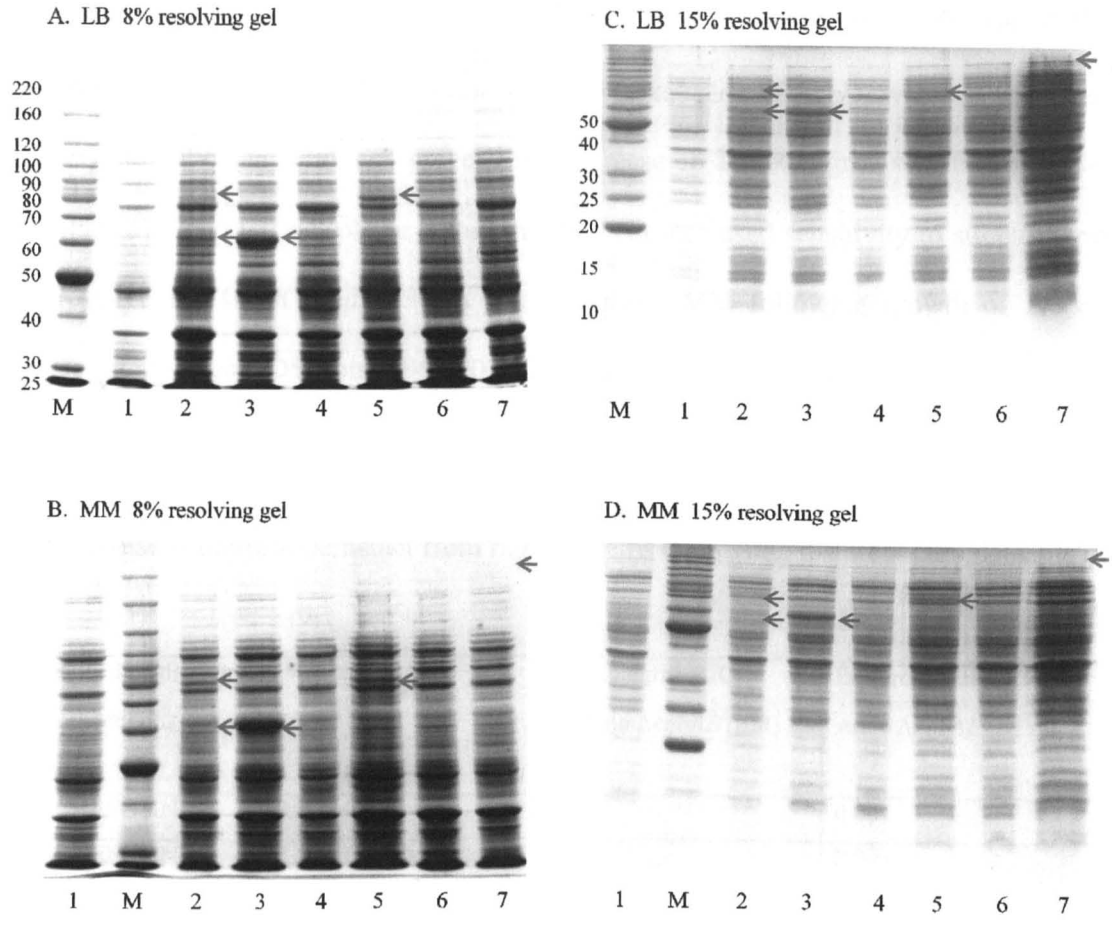


Figure 4.2 SDS-PAGE analysis of *E. coli* strains expressing Eqb proteins

SDS-PAGE gel pictures in panels A to D show soluble protein fractions collected from *E. coli* strains expressing different complements of Eqb proteins following culture in LB broth (panels A and C) or minimal media (MM) (panels B and D). Proteins were separated on 8% (panels A and B) and 15% resolving gels (panels C and D). M: protein marker; lane 1: pre-induction sample from *E. coli* expressing EqbBCDEFGMN proteins; lanes 2-7: 4 h post-induction samples; lane 2: *E. coli* expressing EqbBCDEFGMN proteins; lane 3: *E. coli* expressing EqbBCD proteins; lane 4: *E. coli* expressing EqbMN proteins; lane 5: *E. coli* expressing EqbF protein; lane 6: *E. coli* expressing EqbG protein; lane 7: *E. coli* expressing EqbE protein. Red arrows highlight protein bands corresponding to EqbD (60 kDa, lanes 2 and 3), EqbF (82 kDa, lanes 2 and 5), EqbE (235 kDa, lane 7). Bands corresponding to EqbB (29 kDa), EqbC (26 kDa), EqbM (41 kDa), EqbN (35 kDa) and EqbG (148 kDa) were not observed.

Filter sterilised supernatant from *E. coli* containing all 8 *eqb* CDS and grown in LB increased the sensitivity of strain Δ*eqbAE* to streptonigrin from 1 μM to 0.002 μM, but had no effect on the sensitivity to erythromycin in cross-feeding assays (Table 4.2). Exclusion of *eqbB*, *eqbE*, *eqbF*, *eqbG*, *eqbM* or *eqbN* prevented the associated increased streptonigrin

sensitivity suggesting that all 6 of these proteins are required for reconstitution of the equibactin NRPS in *E. coli* (Table 4.2). However, exclusion of EqbD, which is predicted to acetylate salicylate, did not prevent the increase in streptonigrin sensitivity unless salicylate was also left out of the culture media. An increased sensitivity to streptonigrin on cross-feeding strain Δ*eqbAE* with filter sterilised MM following growth of *E. coli* containing all 8 *eqb* CDS from 0.06 to 0.002 was also observed (Table 4.2).

Table 4.2 Streptonigrin and erythromycin sensitivity in the Δ*eqbAE* strain cross-fed with filter-sterilised culture supernatant from *E. coli* strains expressing different complements of Eqb proteins

Conditioned media from *E. coli* expressing Eqb proteins	Media	Streptonigrin MIC of strain Δ*eqbAE* (μM)	Erythromycin MIC of strain Δ*eqbAE* (μg ml^{-1})
EqbBCDEFGMN	LB	0.06	0.004
EqbBCDEFGMN	LB + 1 mM salicylate	0.002	0.004
EqbCDEFGMN (no EqbB)	LB + 1 mM salicylate	1	0.004
EqbBCEFGMN (no EqbD)	LB + 1 mM salicylate	0.002	0.004
EqbBCEFGMN (no EqbD)	LB	1	0.004
EqbBCDFGMN + pET21a (no EqbE)	LB + 1 mM salicylate	1	0.004
EqbBCDEGMN (no EqbF)	LB + 1 mM salicylate	1	0.004
EqbBCDEFMN (no EqbG)	LB + 1 mM salicylate	1	0.004
EqbBCDEFGN (no EqbM)	LB + 1 mM salicylate	1	0.004
EqbBCDEFGM (no EqbN)	LB + 1 mM salicylate	0.5	0.004
EqbBCDEFGMN	MM + 1 mM salicylate	0.002	0.016
EqbBCDFGMN + pET21a (no EqbE)	MM + 1 mM salicylate	0.06	0.016

Bacteria were grown to stationary phase in LB ± 1 mM salicylate or MM + 1 mM salicylate. MIC refers to the minimum inhibitory concentration of antibiotic required to prevent growth.

However, despite these effects on the streptonigrin sensitivity of strain Δ*eqbAE* in cross-feeding assays, repeated LC-MS analysis of culture supernatants, including conditioned MM, failed to identify equibactin. In addition, no changes were observed in the CAS assay indicative of iron-chelating activity.

4.4.2 *In vitro* reconstitution of equibactin

Recombinant forms of the 8 Eqb proteins (expressed and purified by Carl Robinson, AHT) were mixed with substrates that were predicted to be required by the NRPS catalysts for the biosynthesis of equibactin. Equibactin is proposed to have the same substrates and mechanism of synthesis as yersiniabactin (with the exception of the polyketide component of yersiniabactin synthesis requiring the substrate malonyl-CoA) and reaction conditions accordingly were similar to those used for the *in vitro* reconstitution of yersiniabactin. However, limitations in the quantity and concentration of purified Eqb proteins meant that a lower concentration of biosynthetic proteins was used for the *in vitro* reconstitution of the *eqb* NRPS product(s) (0.2 μM compared to 5 μM used in yersiniabactin biosynthesis) and the experiment was performed only once. Unfortunately, LC-MS analysis (performed by Lijiang Song, Warwick University) failed to identify differences between the test and negative control samples. Neither was a colour change observed in the CAS assay, although reaction mixtures had to be diluted 1 in 10 in order to perform this test without interference from components of the reaction mixture. However, an increased sensitivity to streptonigrin on cross-feeding strain Δ*eqbAE* with filter sterilised reaction mixture (diluted 1 in 4) containing all 8 Eqb proteins from 0.03 to 0.008 (4-fold) was observed.

4.5 Discussion

Reconstitution of the equibactin NRPS in *E. coli* was successfully achieved using three compatable expression plasmids containing *eqbB*, *eqbC*, *eqbD*, *eqbE*, *eqbF*, *eqbG*, *eqbM* and *eqbN*. Production in *E. coli* was enhanced by supplementation of growth media with 1 mM salicylate, was abolished in the absence of *eqbB*, *eqbE*, *eqbF*, *eqbG* and *eqbM* and was reduced in the absence of *eqbN*. Therefore, each of these genes plays a role in the biosynthesis of an active *eqb* NRPS product(s) and the equibactin locus contains all of the genes unique to *S. equi* that are required for the biosynthesis of the NRPS product(s).

The predicted functions of these encoded proteins are discussed in detail in Chapter 3. Clearly, the absence of either of the 2 NRPS proteins (EqbE and EqbG), which contain 7 and 4 of the 17 assigned functional domains, would be expected to severely disrupt the biosynthetic process. In agreement with this result, the Δ*eqbE S. equi* mutant strain was also unable to generate the NRPS product(s) (Chapter 3). *eqbF* encodes a putative thiazoline reductase with some partial homology to YbtU, the thiazoline reductase in *Y. pestis*. *In vitro* reconstitution of yersiniabactin in the absence of YbtU resulted in the production of a molecule that, unlike yersiniabactin, had a thiazoline ring in position 2 (not reduced to thiazolidine) and a thiazole in position one (resulting from spontaneous oxidation) (Miller *et al.*, 2002). The impact of these modifications on yersiniabactin function has not been reported. PchG, the homologous thiazoline reductase of *P. aeruginosa*, introduces a methyl group to the N-atom of the thiazolidine ring in pyochelin (Patel & Walsh, 2001). A similar function catalysed by EqbF in *S. equi* could alter the stereochemistry of the final NRPS product(s) and influence its iron-binding activity as a result.

Interestingly, EqbD, which is predicted to activate salicylate to the adenylate form and initiate the equibactin biosynthetic process, was not necessary for production of the NRPS

product(s) provided that salicylate was added to the growth medium. This result is somewhat surprising and suggests that *E. coli* BL21 (DE3) has an alternative mechanism to activate salicylate. The finding that EqbB was required for the biosynthesis of the NRPS product(s) suggests that this stand alone thioesterase may play a more significant role in the biosynthetic process than is generally observed in other NRPS systems. These type-II thioesterases were initially thought to cleave incorrectly loaded substrates from PCP domains, but more recent work suggests that their role is to hydrolytically remove inactivating acetyl groups from the 4'-phosphopantetheine thiols of PCP domains (originating from the utilization of acetyl-CoA instead of CoA in the post-translational modification of PCP domains) to regenerate active free thiol groups (Challis & Naismith, 2004; Schwarzer *et al.*, 2002). The removal of type-II TEs generally decreases product titres but does not completely abolish product synthesis (Schneider & Marahiel, 1998). Finally, the role played by EqbM and EqbN in the biosynthesis of the NRPS product(s) is unknown. The process used to release the mature NRPS product from the last carrier domain of EqbG is unclear and EqbN, a putative α/β hydrolase may prove to be important here.

In vitro reconstitution of the *eqb* NRPS product(s) appeared successful as evidenced by the 4-fold increase in streptongrin sensitivity of the *Δ*eqbAE strain cross-fed with reaction mixture containing all 8 Eqb proteins. This suggests that further work using higher concentrations of recombinant proteins might yield useful information on the structure of the NRPS product(s) and its mechanism of biosynthesis.

4.6 Conclusions

E. coli has once again proven to be a suitable host for the heterologous production of a foreign non-ribosomal peptide. The ease of genetic manipulation in this host has enabled

a rapid identification of the biosynthetic *eqb* genes necessary for the production of the *eqb* NRPS product(s), at least in *E. coli*. Although these advances have not yet led to the identification of the NRPS product(s), the early success shown using *in vitro* reconstitution suggests that, with improvements, this goal is much closer to being met.

Chapter 5 Quantitative PCR for the detection of *Streptococcus equi*

5.1 Introduction

S. equi is a highly successful pathogen that has plagued the equine population for centuries and contributes to nearly 30% of all infectious diseases reported in horses worldwide (Chanter, 1997). A major facet to this success, as for many bacterial pathogens, is the ability to readily transmit to others and propagate in naïve hosts. This is particularly important to *S. equi*, a bacterium that is unable to survive for long in the environment. Accordingly, measures that interfere with the transmission process contribute considerably to the containment of the disease. Diagnostic assays are clearly important for disease control. Identification of horses that are acutely infected with the pathogen, a process that may simply require confirmation of infection following a display of classical strangles signs, will usually initiate the implementation of movement restrictions and isolation procedures to contain the infection. Of equal importance, the ability of diagnostic tests to detect persistently infected healthy horses that harbor the pathogen can prevent the initiation of future outbreaks. Greater understanding of the contribution that persistently infected carriers have to the spread of strangles has highlighted the need to develop increasingly sensitive assays in order to identify these animals more easily.

For many years, detection of *S. equi* relied solely on morphological and biochemical evaluation, a process that is still in use today. Culture on streptococcal selective blood agar allows the examination of colony characteristics such as size, colour, mucoid appearance and beta-haemolytic activity. The Lancefield capsule agglutination assay is used to distinguish different streptococcal groups. The catalase test is frequently used to differentiate streptococci from staphylococci. Finally, *S. equi* is differentiated from *S. zooepidemicus*, which shares many phenotypic characteristics and is often isolated from the respiratory tract of horses. Discrimination of these 2 subspecies relies on their ability

(or inability) to ferment ribose, sorbitol and lactose. An inability to ferment trehalose separates the *S. equi* and *S. zooepidemicus* from *S. equisimilis*. Culture testing has the advantage of detecting only viable bacteria so that veterinarians are aware of the definitive risk these infected individuals pose to others. Strains collected in this way can also be used for further research. However, this approach suffers a number of disadvantages. Notwithstanding the time consuming, labour and skills intensive nature of such analysis, this methodology also suffers drawbacks of *S. equi* growth inhibition (or masking) by competing bacteria, which ultimately reduces assay sensitivity. In addition, atypical *S. equi* that ferment one or both of lactose and trehalose have been described (Grant *et al.*, 1993).

The era of nucleic acid-based assays has revolutionized the diagnosis of infectious diseases, more recently built on the increasing number of published genome sequences from pathogenic bacteria. Today the major driving forces for developing new diagnostic techniques are reduced hands-on-time and time to diagnosis, as well as increased sensitivity and specificity. These goals are achieved by automating processes and refining detection methods. PCR can detect down to a single molecule of target DNA in a test sample by the amplification of millions of copies lying between oligonucleotide primers, used to direct repeated rounds of synthesis by a DNA polymerase. PCR also has the potential to detect a single target organism amongst millions of others.

The earliest assessment of PCR-mediated *S. equi* identification focused on the 16S-23S rDNA intergenic spacer region but was quickly ruled out due to the inability to design species-specific oligonucleotide primers able to differentiate *S. equi* from *S. zooepidemicus* (Chanter *et al.*, 1997). With this came a discouraging realisation that the identification of *S. equi*-specific target DNA absent from the very closely related but heterogeneous *S. zooepidemicus* population was not going to be an easy task. In 1997, a *S. equi* PCR test

was developed based on the *seM* gene (Timoney & Artiushin, 1997). This test was 3 times more sensitive than culture for detection of *S. equi* in clinical samples (Timoney & Artiushin, 1997) and has been used in the USA as a diagnostic assay for strangles (Holland, 2008). Another version of this *seM*-based test was developed at the AHT using nested PCR (Newton *et al.*, 2000) and evaluated in a number of strangles outbreaks in the UK and Denmark (Gronbaek *et al.*, 2006; Newton *et al.*, 2000). Using NP swab samples from established strangles carriers (n = 61), almost twice as many swabs were positive by PCR (56%) than by culture (30%) and 61% were positive if both results were combined (Newton *et al.*, 2000). Of 12 GP lavage samples collected from established carriers, 76% were positive by PCR and 59% were positive by culture (Newton *et al.*, 2000). The Danish group described a diagnostic sensitivity of 45% for PCR and 18% for culture from nasal swab samples and 80% vs 20% respectively from abscess material (Gronbaek *et al.*, 2006). In addition, of 165 isolates of Lancefield group C identified to species level using standard biochemical tests, 45/45 *S. equi* isolates were positive by PCR, whilst 40 isolates of each species *S. zooepidemicus*, *S. dysgalactiae* and *S. equisimilus* were negative by PCR. This strain collection included epidemiologically independent Danish field isolates (Gronbaek *et al.*, 2006).

These findings were in agreement with previous work, which suggested that SeM is highly homogenous because of the cross-reactivity of sera from a horse convalescent from strangles with a number of different *S. equi* isolates and the lack of variation in HindIII restriction pattern between different *S. equi* isolates on Southern blot analysis using a *seM* gene probe (Galan & Timoney, 1988). However, subsequent work by 2 independent researchers demonstrated differences to the published sequence of the *seM* gene immediately after the region encoding the N-terminal signal sequence (Chanter *et al.*, 2000; Meehan *et al.*, 1998). Even more recently, variation in the 5' region of the *seM* gene has been used as a typing method to differentiate members of the previously

indistinguishable *S. equi* population (Figure 5.1) (Anzai *et al.*, 2005; Kelly *et al.*, 2006; Waller & Jolley, 2007). A publicly available online database at http://pubmlst.org/szooepidemicus/seM/ accessed 12.03.09 shows the identification of a total of 64 alleles of SeM.

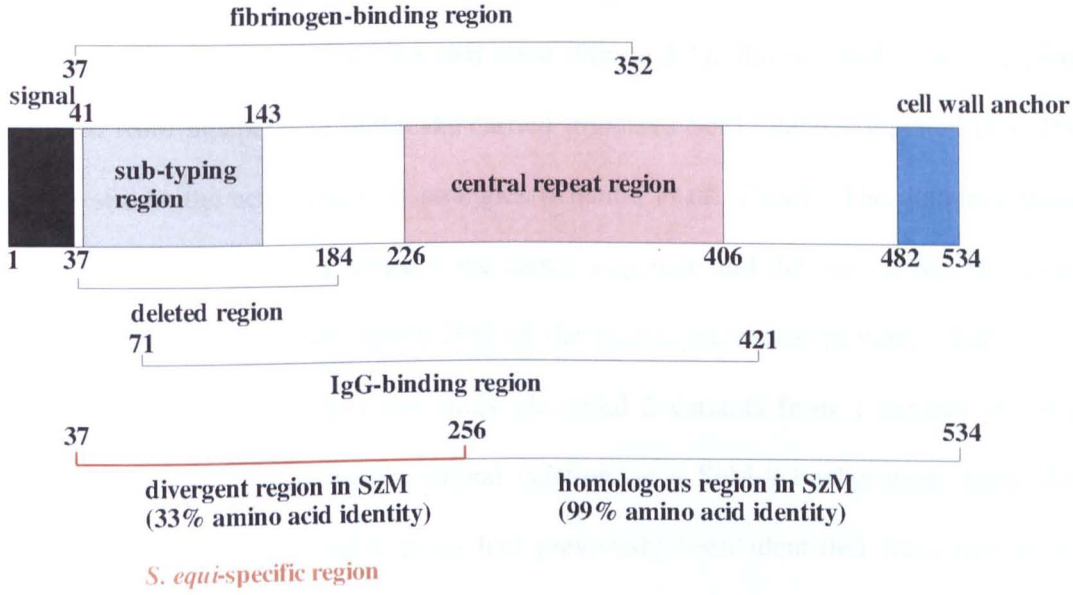


Figure 5.1 Diagram of the SeM protein

Figure shows the functional regions of the *S. equi* SeM anti-phagocytic surface protein. Sequence variation occurs in the sub-typing region highlighted in grey between different strains of *S. equi*. A high ratio of non-synonymous to synonymous substitutions suggests that this sub-typing region is under strong diversifying selection. SzM, the homologous protein in *Sz*H70 shares only 33% amino acid sequence identity in the divergent *S. equi*-specific N terminal region shown in red. In-frame deletions have been identified in the N terminal portions of SeM in some *S. equi* strains isolated from strangles carriers. These fall within the 'deleted region' highlighted in the figure.

The ratio of non-synonymous to synonymous substitutions in this target is extremely high suggesting that the SeM protein is under strong diversifying selection (Waller & Jolley, 2007). This finding is not surprising since SeM is a cell-wall anchored surface protein (Table 2.5, Chapter 2) that is accessible to the host immune system. SeM contributes an important protective antiphagocytic role to *S. equi* (Boschwitz & Timoney, 1994a; Boschwitz & Timoney, 1994b; Chanter, 1994; Meehan *et al.*, 2001; Timoney *et al.*,

1997b). In fact SeM is an immunodominant epitope of *S. equi* and a large proportion of antibody responses are directed to the N-terminus of this protein (Timoney, 1998). Mutation of *seM* over time has been observed during the course of strangles outbreaks (Katy Webb, unpublished data) and circumstantial evidence suggests that SeM sequence diversification aids the persistence of *S. equi* in the GP (Kelly *et al.*, 2006; Waller & Jolley, 2007). Moreover, a small proportion of carriers harbor variants of *S. equi* with N-terminal in-frame regions deleted from the *seM* gene (Figure 5.1). In one study, 24% of carriers sampled from independent outbreaks carried truncated SeM alleles compared to < 1% of horses still in the acute phase of strangles (Chanter *et al.*, 2000). The deletions were in slightly different positions between the signal sequence and the central repeat region of SeM, equivalent to approximately 20% of the mature expressed protein (Chanter *et al.*, 2000). A recent, more extensive study identified 2 variants from 2 carriers in a single outbreak that had a 324 bp N-terminal deletion on a SeM-9 background (Katy Webb, unpublished data). A SeM-9 strain had previously been identified from one of these horses. This was out of 37 carrier isolates (5.4%) identified in the study of 259 *S. equi* isolates (115 horses, 23 independent strangles outbreaks). Such deletions in the *seM* target DNA would lead to false negative results by PCR in this small population of horses. Variation in the primer binding sites might also lead to false negative PCR results. *Sz*H70 encodes a variant of SeM with a unique N-terminus (SzM, Figure 5.1). Therefore, *S. equi*-specific *seM* primers must be designed to the N-terminal variable region of this *S. equi* protein in order to prevent cross-reaction with *S. zooepidemicus* derived DNA.

Overall, these data suggest that more appropriate targets should be investigated in order to reduce the potential risk of infective horses being missed during PCR screening. A multiplex PCR test, based on a *S. equi* subspecies-specific region of the superoxide dismutase gene *sodA* and the *S. equi*-specific mitogenic toxin gene *seeI*, was developed to detect and differentiate *S. equi* and *S. zooepidemicus* subspecies (Alber *et al.*, 2004). A

number of *S. equi* and *S. zooepidemicus* isolates (17 and 26 respectively) were tested using this multiplex PCR. The *sodA* gene was amplified in all isolates; whilst the *seeI* gene was only detected in the *S. equi* isolates (100%). Figure 2.6, Chapter 2 also shows that the *seeI* gene was present in all diverse *S. equi* isolates but absent from the collection of varied *S. zooepidemicus* strains. However, SeeI is a secreted protein of *S. equi* and the authors did propose that a large number of isolates should be screened to ensure that sequence variation in primer binding sites was not occurring (Alber *et al.*, 2004). Allelic variation in the *S. pyogenes* mitogen, SmeZ, highlights the pressure on superantigens to avoid host immune defences (Arcus *et al.*, 2000).

More recently a Swedish group developed a real-time version of this multiplex PCR and correctly assigned the identity of 23/23 *S. equi* isolates and 24/24 *S. zooepidemicus* isolates from a strain collection of the National Veterinary Institute (Baverud *et al.*, 2007). The sensitivity of the real-time PCR test in terms of DNA copy number detected was not established but the test was shown to enhance the detection of *S. equi* by 13% (2/15) compared to conventional biochemical analysis. PCR was performed on DNA extracted from mixed and pure colony material following culture of clinical samples on agar plates. The authors did amplify the *seeI* target in a control strain of *S. pyogenes*, which in a mixed culture with *S. zooepidemicus* could result in a false positive *S. equi* result (Baverud *et al.*, 2007), although *S. pyogenes* is not likely to be detected in equine respiratory samples. The test was not appropriate for the reliable detection of *S. zooepidemicus* since 3 *S. zooepidemicus* isolates with aberrant *sodA* gene sequences were negative by PCR (Baverud *et al.*, 2007).

5.2 Aims

The genome comparison of *Se*4047 and *Sz*H70, aided by gene prevalence screening across a diverse population of typed *S. equi* and *S. zooepidemicus* strains has given an unprecedented opportunity to identify novel diagnostic targets that delineate these subspecies. For the specific detection of *S. equi*, the *eqbE* gene was selected as an ideal DNA target for PCR. *eqbE*, present on ICE*Se2*, was a constant genetic feature of all *S. equi* strains but was absent from the collection of *S. zooepidemicus* strains tested (Chapters 2 and 3). In addition, no close homologues of *eqbE* were identified in any other organisms in the public databases. The work described in this Chapter set out to further assess the suitability of the *eqbE* target for the detection of *S. equi* by PCR. I also intended to make use of new technological advances and assess the suitability of the *eqbE* target for quantitative real-time PCR. Since *eqbE* PCRs were developed for Chapters 2 and 3, the specific objectives for this Chapter were as follows:-

1. To sequence a diverse set of *S. equi* strains in order to examine sequence variability in the *eqbE* target.
2. To develop a real-time quantitative *eqbE*-based PCR test and assess the sensitivity of this assay.

5.3 Methods

5.3.1 DNA extraction

To extract a high yield and quality of genomic DNA from *Se*4047 to serve as a stock for the preparation of PCR standard controls, the following protocol was used. A 10 ml overnight starter culture of *Se*4047 in THB was subcultured into 200 ml fresh THB supplemented with 30 $\mu g ml^{-1}$ hyaluronidase and grown overnight at 37°C with 5% CO_2. Cells were harvested (5,000 g, 10 min, room temperature) and resuspended in 20 ml 0.1 M

EDTA, 0.15 M NaCl, pH 8.5. Lysozyme (10 mg) and mutanolysin (5 kU) were added and the suspension was incubated without shaking at 37°C for 2 h for lysis to occur. 200 µl RNase solution (R6148, Sigma) was added and the suspension was incubated for a further 10 min at room temperature. 1 ml of 10% SDS was then added, followed by 4.2 ml 5 M sodium perchlorate and 50 ml chloroform/octanol (24:1 volume:volume), with thorough mixing between each addition. The mixture was then mixed thoroughly for 15 min at room temperature and centrifuged (15,000 g, 20 min, 4°C). Following centrifugation, the top aqueous layer was removed and added gently to a beaker containing 2.5 volumes of ethanol to create 2 phases. DNA was recovered by spooling on a glass hook, washed in 70% ethanol (5 ml) and resuspended in 1 ml 10 mM Tris-HCl, pH 9.0, 0.5 mM EDTA. The quantity and purity of DNA was determined by optical density measurements at 260 and 280 nm using a NanoDrop^{R} ND1000 spectrophotometer (NanoDrop Technologies). Agarose gel electrophoresis was also used to assess the quality of the DNA extracted.

DNA template copy number was determined using the formula:-

$$\text{Number of copies} = \frac{\text{amount of DNA (ng)} \times 6.022 \times 10^{23}}{\text{length of DNA (bp)} \times 1 \times 10^{9} \times 650}$$

This calculation assumes that the average weight of a bp is 650 Daltons. The length of the *S. equi* chromosome is 2,253,793 bp.

All other DNA extractions were performed on colony material scraped from agar plates or collected by centrifugation (1 ml) from overnight cultures grown in THB at 37°C with 5% CO_2. DNA was extracted using the GenElute Bacterial Genomic DNA kit (Sigma) according to manufacturer's instructions.

5.3.2 Real-time quantitative PCR

Table 5.1 Oligonucleotides used for *eqbE* endpoint-PCR, *eqbE* quantitative PCR and sequencing across the *eqbE* PCR target sites

Primer/ probe	Sequence (5' to 3')	Purpose
EqbE f	AAGATATAGCAGCATCGTATCG	QPCR and sequencing
EqbE r	TCTAAATCTCTATTAAATAGCGGTATATTG	QPCR
Equidetectin	TCTaTGgTTcTTcTAACTGCCTATGC	Dual labelled hybridization probe for QPCR
2f	GGGTTGCCATGCATATCTTG	*eqbE* endpoint-PCR
2r	TCCGGCTGTTTCCTTAATGG	
ZM435	CCGAATTTGTCCAAGTGGTATG	Sequencing
ZM436	GCACTCCGTTATACTCACTG	
ZM437	TTTGCTAGTGCTACTCCTGC	

Locked nucleic acids (LNAs) are shown in undercase.

Probe and primer sequences were designed by Sigma using Beacon Designer software (Table 5.1). The DNA probe was dual labelled with 5' 6-FAM and 3' BHQ-1 (Black Hole Quencher-1) and contained LNAs to improve probe thermostability. Suitable primer and probe concentrations for QPCR were established following optimization experiments. All reactions were performed in duplicate or triplicate. The optimized 20 µl reaction contained test DNA sample (2 µl) or standard *Se*4047 DNA of a known template copy number, 0.3 µM forward and reverse primers (Sigma), 150 nM DNA probe (Sigma) and 1x ABsolute Blue QPCR mix (Abgene). Thermocycling was performed at 95°C for 15 min, followed by 50 cycles of 95°C for 15 s, and 60°C for 30 s. Fluoresence was read during the annealing/extension step of each cycle. Positive, negative and no-template controls were included. QPCR was performed in Techne Thermofast 96 non-skirted low-profile plates sealed with clear seal diamond optically clear heat sealing film (Thermo Scientific) in a Techne Quantica instrument and data analysed using Quansoft software (Techne). Crossing point (Cp) was determined for each sample or standard using the default settings of the Quansoft programme. Fluorescence was plotted on a log scale against cycle number (Figure 5.2). A noise threshold was set 4 standard deviations (SD) above background

fluorescence, which was calculated as the average fluorescence between cycles 3 and 10. A crossing threshold was set 10 SD above background fluorescence. Cp was determined using 2 fit points on the curve above the crossing threshold to calculate by linear regression the best-fit gradient. The intersection of this line through the crossing threshold was defined as the Cp in terms of cycle number. Standard curves (Cp vs log gene copy number) were generated to assess PCR efficiency and calculate the number of template DNA copies in an unknown sample.

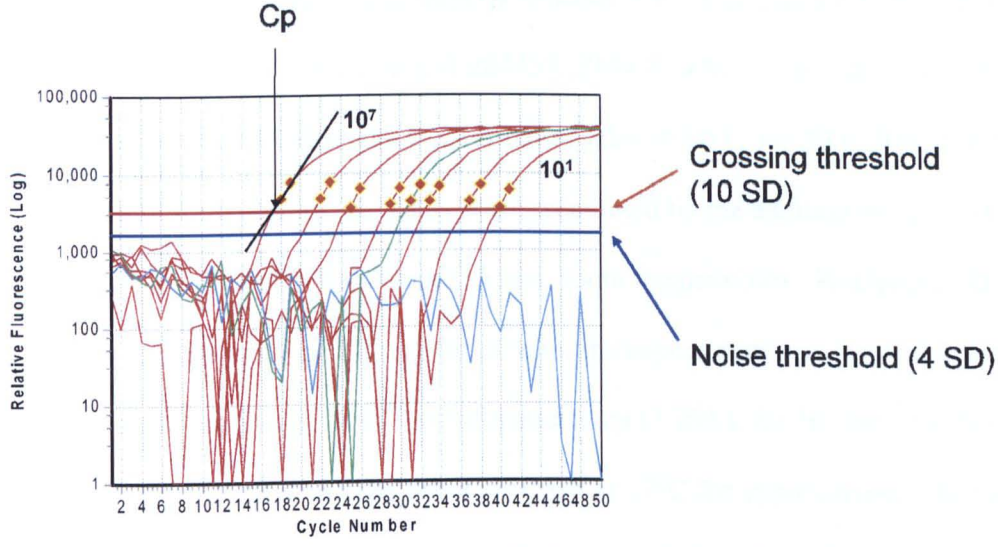


Figure 5.2 Calculation of crossing point (Cp)

Relative fluorescence is plotted on a logarithmic scale against cycle number. Red lines represent signal from standard controls, the green line is an unknown sample and the blue line is a no-template control. Crossing threshold and noise threshold are marked at 10 SD and 4 SD above background fluorescence, respectively. 2 fit points are marked on each curve above the crossing threshold and a line of best fit to these points is marked. Cp is determined where this line intersects the crossing threshold.

QPCR was also performed using the above conditions without probe and with the ABsolute QPCR SYBR Green mix (Abgene) instead of the ABsolute Blue QPCR mix (Abgene). Dissociation curves were analysed following a final ramp step from 60°C to 90°C with reads at 0.5°C increments. This SYBR Green QPCR was performed to check that the primers were not mispriming and generating non-specific amplification products.

5.3.3 Sequence analysis

A 1,063 bp region of the *eqbE* gene that included both endpoint-PCR and QPCR target sites was sequenced in 26 strains of *S. equi* (including diverse SeM and MLST types, Table 5.2) to check for sequence variability. PCR was performed using Phusion DNA polymerase (NEB) according to the manufacturer's instructions using the ZM435 and ZM436 primers (Table 5.1). The amplified product was purified using the QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. Sequencing was performed using BigDye® Terminator Cycle Sequencing (Applied Biosystems). A 15 µl sequencing reaction contained 2 µl BigDye reaction mix, 2 µl dilution buffer (0.4 M Tris pH 9.0, 10 mM $MgCl_2$), 1.6 µl 10 µM ZM435, ZM436, ZM437 or EqbE f primer (Table 5.1) and 1 µl purified PCR product. 25 thermocycles of 96°C for 30 s, 50°C for 15 s and 60°C for 4 min were performed. DNA was precipitated by the addition of 2µl 3 M sodium acetate, pH 4.6 and 50µl 95% ethanol (15 min, room temperature). Precipitated DNA was pelleted by centrifugation at 2,254 g for 30 min, the supernatant was discarded and 180 µl of 70% ethanol was added. A 2nd centrifugation step (2,254 g for 10 min) was performed, the supernatant discarded and the pellet air dried at 37°C for approximately 30 min. The DNA was finally resuspended in 10 µl HiDi formamide (Applied Biosystems), heated to 95°C for 2 min and analysed with an ABI 3100 DNA sequencer. SeqMan 5.03 DNAstar software (Lasergene) was used to analyse sequence reads.

Table 5.2 Identity of *S. equi* isolates screened for sequence variability in the *eqbE* gene target

SeM allele	ST	Source	Isolate ID	Disease/ Source	Animal	Year
2	179	Ireland	SA	Strangles	Horse	1999
2	179	Canada	303	Strangles	Horse	1999
2	179	New York	CF32	Strangles	Horse	1981
3	179	Hampshire	*Se*4047	Strangles	Horse	1990
5	179	Leicestershire	B23 7325	Strangles	Horse	2003
5	179	Leicestershire	7326	Strangles	Horse	2003
6	179	Hertfordshire	8229	Strangles	Dog	2004
6	179	Suffolk	7060	Strangles	Horse	2003
6	179	Kent	1351	Strangles	Horse	2004
6	179	Essex	1350	Strangles	Horse	2003
6	179	Norfolk	JKS 044	Strangles carrier	Horse	2006
6	151	Gwent	7329	Strangles	Horse	2006
7	179	Suffolk	1931	Strangles	Horse	2004
9	179	Hampshire	7344	Strangles	Horse	2003
9	179	Essex	JKS 559	Strangles	Horse	2007
10	179	Hampshire	7140	Strangles	Horse	2003
10	179	Shetland	7171	Strangles	Horse	2003
11	179	Suffolk	7364	Strangles	Horse	2003
12	179	Suffolk	3154	Strangles	Horse	2004
13	179	Suffolk	3155	Strangles	Horse	2004
14	179	Suffolk	3156	Strangles	Horse	2004
15	179	Australia	181063	Strangles	Horse	1999
16	179	Lincolnshire	JKS 063	Strangles	Horse	2006
44	179	Lincolnshire	JKS 225	Strangles	Horse	2006
47	179	Norfolk	JKS 043	Strangles	Horse	2006
52	151	UK	Mo	Strangles brain abscess	Horse	2008

5.4 Results

5.4.1 A conserved region in *Streptococcus equi* strains

No sequence variability occurred in the 1,063 bp region of the *eqbE* gene amplified by ZM435 and ZM436 primers in 26 *S. equi* isolates. Primer sets 2f/2r and EqbE f/ EqbE r were used for the *eqbE*-based end-point PCR (Chapter 3) and QPCR, respectively. These primers hybridize within the DNA fragment amplified by ZM435 and ZM436 in a region of the *eqbE* NRPS gene that encodes a putative condensation/heterocylization domain. The diverse set of isolates examined was representative of different SeM-types found around the world (Table 5.2 and Figure 1.4, Chapter one) and included the 2 most prevalent variants SeM-9 and SeM-6 (Kelly *et al.*, 2006), (Katy Webb, unpublished data). *S. equi* strains are mostly MLST ST-179 but 2 isolates have been identified with ST-151, which is a single locus variant of ST-179. These were included in the analysis; one of which was recovered from a brain abscess, which is an unusual manisfestation of *S. equi* infection. One isolate tested was recovered from the first documented case of canine strangles (Ladlow *et al.*, 2006) and another was isolated from a persistently infected strangles carrier (Table 5.2).

5.4.2 Development of a real time quantitative PCR test for the detection of *Streptococcus equi*

A dual labelled-probe based QPCR method was developed for the detection of *S. equi*. Probe and primer sequences specific for the *eqbE* gene were designed for QPCR using Beacon Designer software by Sigma (Table 5.1). The DNA probe was labelled with 5' 6-FAM and 3' BHQ-1 (Black Hole Quencher-1) and contained LNAs. A LNA nucleotide has a modified ribose moiety with an extra bridge connecting the 2' and 4' carbons. LNAs were included to increase the strength of the interaction between the probe and target sequence in a region of DNA that was rich in A and T nucleotides. This aimed to ensure probe hybridisation prior to primer annealing so that extension by DNA polymerase

resulted in efficient exonucleolytic cleavage of the bound probe, release of the 6-FAM dye and a resultant increase in fluorescence as the 6-FAM moves away from the quencher dye. A higher melt temperature also promotes the specificity and sensitivity of probe hybridization to the target DNA.

Suitable primer and probe concentrations for the QPCR were established following optimization experiments. Primer titrations using 50 nM, 300 nM and 900 nM concentrations of each primer were assessed to establish the optimum primer concentrations for the amplification of a 130 bp *eqbE* DNA fragment from a *Se*4047 genomic DNA sample. 300 nM of EqbE f and 300 nM EqbE r gave the lowest Cp reading without introducing non-specific amplication (Figure 5.3A). Primer concentrations of 300/900, 900/300 and 900/900 gave lower Cp values but non-specific products were sometimes amplified late in these PCR reactions. Probe concentrations of 50 nM, 100 nM, 150 nM, 200 nM, 250 nM and 300 nM were evaluated. 150 nM was the minimal concentration of probe required to give the lowest Cp value (Figure 5.3B). The highest PCR efficiency was also achieved using 150 nM probe compared to 50 nM, 100 nM and 250 nM probe (Table 5.3). Optimum PCR conditions therefore had 300 nM of forward and reverse primer and 150 nM of DNA probe. Using these conditions a standard curve was generated from *Se*4047 genomic DNA template controls ranging in concentration from 10^7 copies to 10 copies through 10-fold serial dilutions (Figure 5.4). These standards fell within the linear working range of the assay ($R^2 = 0.998$), which had a PCR efficiency of 93%. A good QPCR assay is usually expected to have efficiency within the range 90% - 110%. One copy of template DNA could be detected using this assay, but not in a consistent fashion (not shown). Generally no fluorescent signals were recorded from negative or no-template controls but occasionally a signal was observed very late in the PCR reaction at a cycle number equivalent to less than 10 copies of DNA, possibly as a result of cross contamination. A SYBR Green assay showed that only one product with a

melt temperature of 77.5°C was being amplified from the EqbE f and r primers at 300 nM concentrations (Figure 5.5).

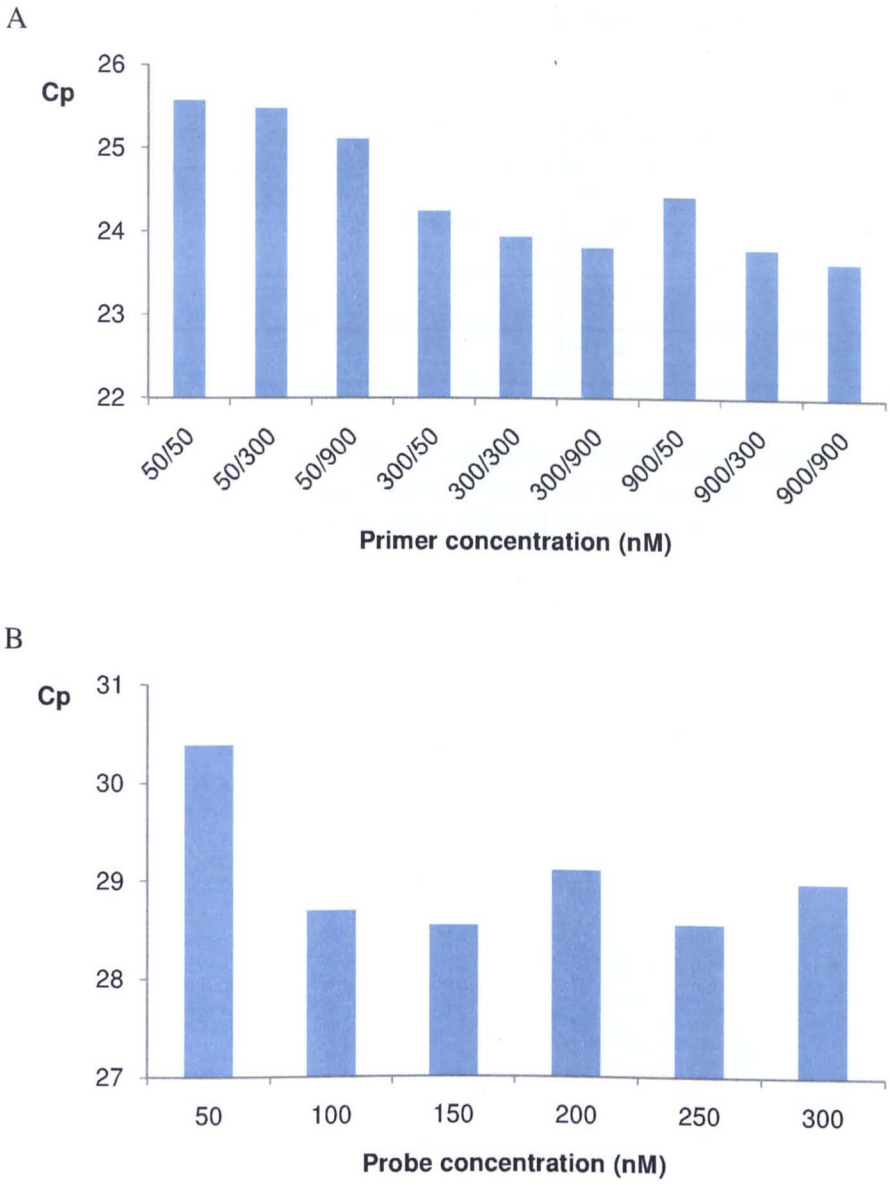


Figure 5.3 Optimisation of primer and probe concentrations for QPCR

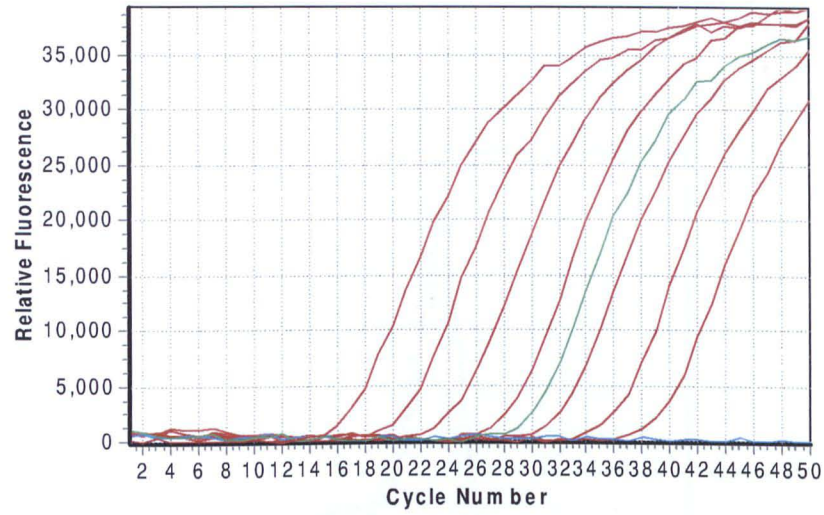
(A) Primer concentrations of 300/900, 900/300 and 900/900 (nM/nM) gave the lowest Cp values but non-specific amplification products were sometimes amplified late in these PCR reactions. 300/300 (nM/nM) was therefore picked as an optimum concentration of forward and reverse primer. (B) 150 nM probe gave the lowest Cp value.

Table 5.3 Influence of probe concentration on QPCR efficiency

Equidectin probe concentration (nM)	QPCR efficiency (%)
50	91.9
100	92.6
150	93.4
250	91.4

QPCR on 7 *Se*4047 genomic DNA standards ranging in concentration from 10^7 to 10 DNA copies through 10-fold serial dilutions was used to construct a standard curve and calculate QPCR efficiency for reactions using 300 nM forward and reverse primer and varying concentrations of probe.

A



B

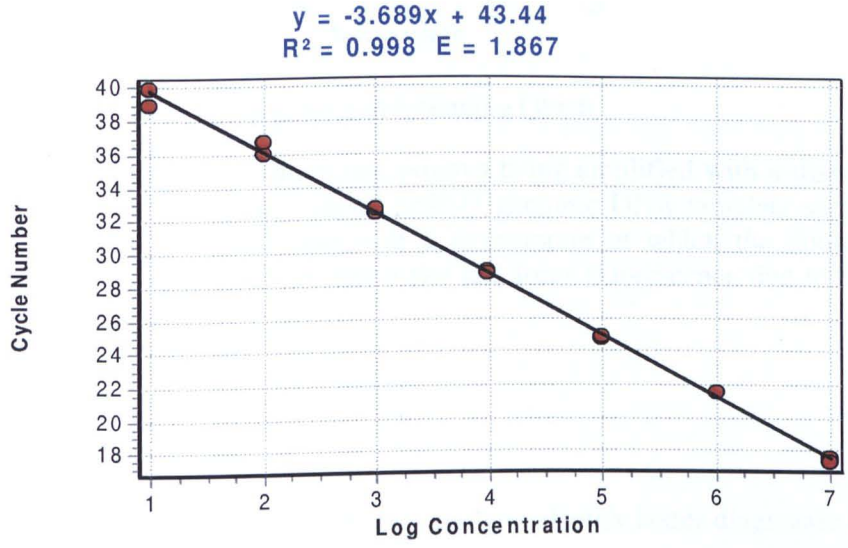


Figure 5.4 *eqbE* QPCR standard curve

(A) PCR cycle number plotted against relative fluorescence shows the increase in fluorescence as the PCR progresses. Red lines show *Se*4047 genomic DNA template standard controls ranging from 10^7 copies to 10 copies in 10-fold serial dilutions. Higher concentrations of template DNA are amplified earlier in the PCR reaction. The green line represents a sample of previously unknown *S. equi* DNA concentration. No-template controls are shown in blue. (B) Log template concentration is plotted against Cp (cycle number) to generate a standard curve. The equation of the line shown above the graph is used to calculate the quantity of *S. equi* DNA in unknown samples from their Cp value. The R^2 value of 0.998 shows a good fit to the straight line (for a perfect fit $R^2 = 1$) and the linear range covers 10^7 to 10 copies of template DNA. An E value of 1.867 demonstrates good PCR efficiency. An efficiency of 2 is achieved from an exact doubling of PCR product with each cycle. Similarly, a gradient of -3.32 represents 100% doubling efficiency.

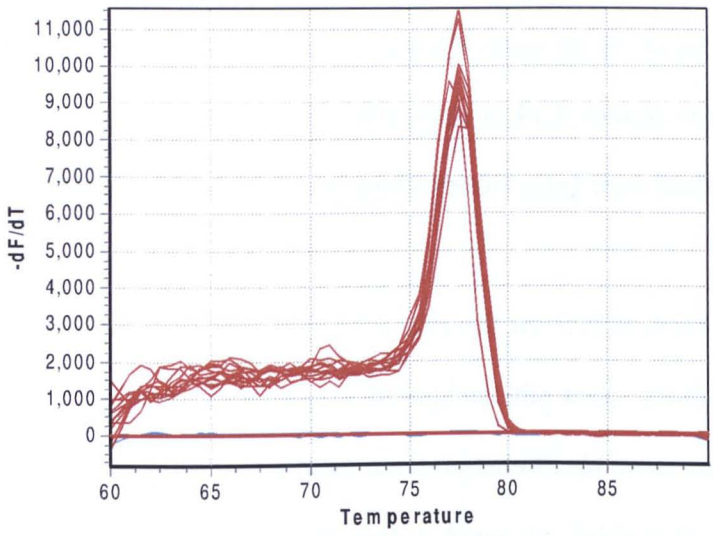


Figure 5.5 Melt curve generated following QPCR

A SYBR green assay shows one product being amplified with a dissociation temperature of 77.5°C from standard control *Se*4047 genomic DNA template samples. The negative first derivative (-dF/dT) peaks at a temperature at which the double stranded 130 bp fragment of *eqbE* PCR product melts and loses fluorescence due to loss of SYBR green intercalation.

5.5 Discussion

Improving technology provides new tools to identify better diagnostic targets and to detect those targets more effectively. Comparative genomic analysis revealed a novel genetic locus in *S. equi*, an NRPS system encoded on ICE*Se2*, which may have contributed to the evolution of this important veterinary pathogen through increasing its ability to acquire iron (Chapters 2 and 3). This locus also offers an ideal target for PCR detection of *S. equi*. Although the product of this NRPS system is secreted, the *eqbE*-encoded enzyme involved in its biosynthesis has a predicted location in the intracellular compartment of the cell. Consequently, EqbE is not likely to be subjected to the high levels of diversifying selection suffered by other surface exposed antigens, such as SeM (Kelly *et al.*, 2006). The *eqbE* gene is also unlikely to vary in response to immune selection imposed on the secreted

peptide. Non-ribosomal peptides may be too small to elicit an immune response and, unlike ribosomally synthesised proteins, have little opportunity to vary in structure and still retain their activity. The diverse strains of *S. equi* examined here differed in their *seM* gene sequences, but had identical *eqbE* genes (in the target region examined). As mutation in primer binding sites generally leads to PCR failure, the variation in *seM* sequence has the potential to cause false negative results using tests that target this gene.

In addition, a variant of SeM is encoded by *Sz*H70 (SzM, Chapter 2) and other forms of SzM are likely to be spread throughout the heterogenous *S. zooepidemicus* population. Primers designed to the N-terminal region of *seM* were used to screen some of the diverse *S. zooepidemicus* strains detailed in Appendix Table A.4. Although this data has not been reported in this thesis, some isolates of *S. zooepidemicus* were positive by PCR suggesting that a small proportion of the *S. zooepidemicus* population encode a close SeM homologue. A risk of false positive reporting is therefore possible using *seM*-based PCR tests. Conversely, *eqbE* was not detected in any of the diverse *S. zooepidemicus* isolates by PCR (Chapters 2 and 3) and *eqbN* was absent from *S. zooepidemicus* isolates despite using low stringency conditions for detection by Southern blot (Chapter 2), suggesting that this locus is specific to *S. equi*. Moreover, the predicted amino acid sequence of EqbE had no close homologues in the public databases (Chapter 3). *eqbE* as a PCR target therefore offers advantages over the *seeI* PCR target that can be detected in some strains of *S. pyogenes* (Alber *et al.*, 2004; Baverud *et al.*, 2007). NRPS systems are rare in streptococci and potential siderophore biosynthetic systems have not yet been detected in streptococci (Donadio *et al.*, 2007) other than *S. equi*.

An end-point PCR, developed in Chapter 3, that amplifies part of the *eqbE* gene was passed on to the AHT diagnostic team in 2006 for further validation and was subsequently commercialised for strangles testing. The *eqbE*-PCR had an approximate 5-fold improved

sensitivity compared to a *seM*-based PCR test being used at the time and also performed better on clinical isolates (Olivia Wahlan, unpublished data).

The *eqbE*-PCR test was later redesigned to allow analysis by QPCR. A QPCR test with 93% efficiency was developed, which could detect as few as 10 copies of *S. equi* genomic DNA (Figure 5.4). The QPCR was found to be more sensitive than the end-point PCR, which could detect 50 copies of template DNA. In addition, the probe hybridisation strategy used for detection of QPCR amplification confers an additional level of specificity to the test as the probe and both primers are required to anneal to the target sequence in order for the QPCR to work. The QPCR method was passed on to the AHT diagnostic team for further validation and is now in operation as a commercial test. A patent application has been submitted to permit licensing of this test to other diagnostic laboratories around the world.

QPCR offers many advantages over the classical end-point PCR test. One important benefit is the quantification of the amount of *S. equi* DNA in the sample being tested. For example, a high level of *S. equi* detected in a clinical wash sample may leave a practitioner in less doubt that non-viable bacteria or cross-contamination of equipment during sampling could have contributed to a positive result. Customer confidence usually leads to greater usage, which has obvious advantages to the control and even eradication of strangles. Detection of low levels of *S. equi* should not lead to the assumption that either of these scenarios has occurred. Intermittent NP shedding is a feature of *S. equi* infection and individuals carrying the pathogen in the GP may well shed fewer bacteria. Instead, these animals should be investigated further with repeat testing and/or GP endoscopy. Understanding how to use available diagnostic tools to their best advantage is an important aspect of disease control. Access to quantitative tools may also help to assess recovery from infection, including GP infections that often require lengthy antibiotic therapy. Other

advantages offered by the QPCR test include reduced, safer user handling with improved turnaround times and lower contamination risks.

5.6 Conclusions

Comparative genomic analysis has led to the identification of an excellent *S. equi* specific target for diagnostic PCR. The PCRs developed based on this DNA target offer improved sensitivity and potentially improved specificity with a reduced risk of false negative or positive results being reported. Already commercially available in the UK, the wider application of these tests may offer significant improvements to the diagnosis and control of strangles, particularly with the new quantitative aspect offered. In the future it may be possible with new technologies to translate this test onto a platform suitable for point of care diagnosis.

Chapter 6 Overall conclusions and future work

The genome comparison between *Se*4047 and *Sz*H70 has provided some explanations for the differences in host range and pathogenesis of these closely related organisms. In evolving from a versatile *S. zooepidemicus* ancestor to a highly specialised obligate equine pathogen, major adaptive features of the *S. equi* genome include the gain of MGEs, a loss of ancestral capabilities, and a change in the organisation or sequence of a number of genes (Holden, 2009).

Gene prevalence screening by PCR across diverse members of each subspecies helped to clarify genetic features that are unique to *S. equi* and identified members of the *S. zooepidemicus* population that share common features with *S. equi* with the potential to promote virulence. Quick, 'next-generation', sequencing technologies will allow many more *S. equi* and *S. zooepidemicus* genomes to be sequenced so that genetic changes that have contributed to the emergence of *S. equi* will be better defined and an order to these events described. This may highlight the importance of genetic factors identified in this study, but not yet analysed further. Future sequencing should also help to define the extent of heterogeneity in the largely clonal *S. equi* population, which will be critical to future epidemiological analysis, vaccine development and the identification of changes that may occur as a result of host-pathogen interactions, for example, during persistence in the GP. Where vaccines are desired to prevent disease associated with *S. zooepidemicus* infection, more isolates can be sequenced to identify linked fitness genes that might prove suitable for vaccination.

6.1 The contribution of prophage exotoxins to pathogenicity or survival fitness of *S. equi* and strains of *S. zooepidemicus*

In this study, *S. equi* was shown to have acquired 4 prophage, encoding a combination of 4 superantigens and a phospholipase A_2 toxin. A small number of *S. zooepidemicus* isolates also contained *speL* and *speM*, and further strains were predicted to encode novel superantigens. In addition, a significant proportion of the *S. zooepidemicus* population contained the phospholipase A_2 gene, introduced into *Se*4047 by φSeq2, but possibly acquired on a different prophage backbone in *S. zooepidemicus*. Superantigens and phospholipase A_2 toxins are hypothesised to corrupt the equine immune response and facilitate the pathogenesis or survival fitness of *S. equi* and strains of *S. zooepidemicus*. The ability of these toxins to misdirect the equine immune system and thereby disrupt efficacious responses could be addressed using *in vitro* assays that quantify lymphocyte proliferation, cytokine responses, T cell anergy and antigen presentation. Early progress has already been made in this area, with SeeI, SeeL and SeeM shown to induce lymphocyte proliferation and the synthesis of interferon-γ ((Artiushin *et al.*, 2002), Romain Paillot, unpublished data). The contribution of exotoxins to pathogenicity could be established using *in vivo* models of disease and allelic replacement strains lacking specific toxin genes.

Based on previous work carried out with GAS, superantigens and/or phospholipase A_2 enzymes are hypothesised to enhance specifically the adherence, colonisation, host cell cytotoxicity and/or dissemination exhibited by *S. equi* and certain strains of *S. zooepidemicus*. These characteristics could be assessed during *in vivo* infection and explored using *in vitro* assays. Human cell lines have been utilised to investigate the contribution of phospholipase A_2 (SlaA) to GAS adherence and killing of epithelial cells, and similar systems could be established using equine cell lines. Alternatively, an already

established air-interface model of the equine respiratory tract (Hamilton *et al.*, 2006) could be used to quantify these attributes at different time points following infection of organ cultures by wild-type and mutant strains of *S. equi* or *S. zooepidemicus*. This model may be more appropriate for studying *S. zooepidemicus* infection, as *S. equi* appears unadapted to colonise the upper respiratory tract for more than a few hours (Timoney & Kumar, 2008). However, it may be possible to identify differences between different *S. equi* strains, particularly early adherence and invasive potential using electron microscopy and bacterial counting techniques. Host cell cytoxicity could be monitored potentially by quantifying levels of lactate dehydrogenase released by tissue (Sitkiewicz *et al.*, 2006), although this has not been attempted previously using the air interface model. Of course such *in vitro* models lack a humoral immune response, which may negate the role played by these exotoxins during the infection of organ cultures. Since *Se*4047 SlaA is nearly identical to the orthologous protein in GAS, this protein is hypothesised to have similar enzymatic activity against phospholipids (Nagiec *et al.*, 2004), which could be confirmed using the same *in vitro* assay previously described (Nagiec *et al.*, 2004) and any differences in phospholipase A_2 activity elicited by SlaA and SlaB could be established.

6.1.1 Does host variation impact on superantigen-mediated pathogenicity in *S. equi*?

Whilst superantigens are hypothesised to contribute to the increased pathogenicity of *S. equi*, their full contribution to virulence may be dependent on individual host factors. Polymorphisms in the genetic background of the host, for example in MHC class II molecules, have been shown to result in quantitative and qualitative differences in cytokine responses to certain GAS superantigens (Fraser & Proft, 2008). Individual horses may therefore respond differently to the various *S. equi* toxins and reflect this in the severity of disease shown (including disease complications) or their carrier status. Future experimentation should allow for potential differences in the specificities of some

superantigens to different MHC class II molecules and investigate a possible link to disease severity. Similarly, the production of neutralising anti-toxin antibodies and associations to severity of disease could be examined in individual horses, as an absence of neutralising antibodies has been highlighted as a significant risk factor for susceptibility to toxin shock syndrome associated with *S. aureus* or *S. pyogenes* infection in humans (Fraser & Proft, 2008). Early experiments have already detected specific responses to *S. equi* superantigens in convalescent sera from strangles cases (Nicola Butcher, unpublished data) and serum samples have been shown to neutralise SeeI mitogenic activity (Artiushin *et al.*, 2002), (Romain Paillot, unpublished data). However, these initial experiments detected an inconsistent production of neutralising antibodies against SeeL and SeeM (Romain Paillot, unpublished data) between individuals, which is a finding that deserves further investigation.

6.1.2 Future vaccines based on prophage exotoxins

Protection against *S. equi* infection is hypothesised to require the targeting of *S. equi*-specific virulence factors, such as superantigens highlighted in this study, as horses that are regularly exposed to *S. zooepidemicus* antigens remain susceptible to strangles. Immunisation with regions of superantigen proteins have conferred protection against *S. aureus* disease in animal models (Ulrich *et al.*, 1998). Vaccination with partial regions of recombinant phospholipase A_2 or superantigen toxin(s) or full length proteins that have been mutated to prevent their toxic effects could therefore be assessed for their ability to produce neutralising antibodies and to protect against *S. equi* (and in some cases *S. zooepidemicus*) infection using *in vivo* models. As superantigens are also hypothesised to interfere with the generation of specific immune responses, similar mutations could be introduced into the toxin genes of *S. equi* live attenuated vaccine strains in an attempt to

improve the induction of constructive immune responses, whilst still presenting epitopes for the generation of specific immunity to these toxin proteins.

6.1.3 A common phage pool and novel superantigens in strains of *S. zooepidemicus*

This research has highlighted a common phage pool shared relatively recently by both group A and group C streptococci, suggesting that the evolution of these important human and veterinary pathogens can not be considered in isolation. A number of *S. zooepidemicus* isolates had mitogenic activity (Holden, 2009), but this study showed that they were negative by PCR for superantigen genes found in *S. equi*. Future sequencing is therefore required to identify potentially novel superantigens encoded by these strains and any prophage that may contribute to their mobilisation. The ability of prophage to revert to lytic infection and the properties of new superantigens could then be assessed.

6.2 Defence against phage infection

Gene loss has been revealed during this study as a major feature of *S. equi* evolution, which in the case of the CRISPR locus and possible competence associated defence systems, may have contributed to the formation of polylysogeny in *S. equi* and the gain of multiple phage cargo. A CRISPR-associated gene was identified in most strains of *S. zooepidemicus* examined, including many of the isolates with superantigens, *slaA* or mitogenic activity. This suggests that CRISPRs generally help to maintain genome stability and prevent deleterious effects of phage infection, although situations are also likely to arise whereby phage transduction and the gain of an associated exotoxin confer a selective advantage that promotes strain survival. Prophage may be subsequently elimited from the chromosome as has presumably occurred with *slaB*. Further sequencing of *S. zooepidemicus* genomes might highlight the contribution of CRISPR or competence loci to phage defence and could be studied in more detail using phage infection studies.

6.3 Gene loss in *S. equi*: genetic drift or positive selection?

An accumulation of pseudogenes and the proliferation of IS elements suggests that *S. equi* passed through a population bottleneck during its transition into a new specialised niche. Gene loss, which is most evident in catabolic metabolism, transport and the cell envelope, may have arisen largely from genetic drift rather than a response to environmental pressures. In the case of sugar metabolism, the gene decay responsible for the biochemical differentiation of *S. equi* from *S. zooepidemicus* has been established for the first time and a requirement to include lactose fermentation in diagnostic *S. equi* detection has been highlighted and subsequently put into practise in the AHT diagnostic laboratory. However, some gene inactivations may have been positively selected for. The truncation of the fibronectin-binding protein, Fne, tallies with this pathogen's reduced fibronectin-binding capacity, which has previously been hypothesised to be an important factor behind the invasive nature of *S. equi* infection (Lannergard *et al.*, 2005; Lindmark & Guss, 1999; Lindmark *et al.*, 1999). This study has revealed a number of *S. equi* surface proteins that have been truncated or deleted and I hypothesise that further adhesion mechanisms have been lost in *S. equi*, which contribute to its decreased colonisation potential and may in some cases have increased its pathogenicity. Secreted adhesins that have lost their cell wall anchor may induce changes in the host cells through interactions with host integrins and ECM components, as observed with Fne (Liden *et al.*, 2008). This small subset of decayed, secreted surface proteins with retained N-terminal ligand binding moieties should probably not be ruled out as potential virulence factors and could be investigated further.

6.4 The contribution of surface proteins lost by *S. equi* to survival fitness or pathogenicity in strains of *S. zooepidemicus*

Surface proteins lost during the evolution of *S. equi*, are likely to contribute to the ability of *S. zooepidemicus* strains to colonise different hosts or certain tissue types and in some

cases cause opportunistic disease. 2 loci in particular are hypothesised to promote properties such as bacterial aggregation, biofilm formation, host cell adhesion and/or invasion (Abbot *et al.*, 2007; Dramsi *et al.*, 2006; Gaillard *et al.*, 1991; Krishnan *et al.*, 2007; Maisey *et al.*, 2007; Mandlik *et al.*, 2007; Mandlik *et al.*, 2008; Manetti *et al.*, 2007; Nelson *et al.*, 2007; Pezzicoli *et al.*, 2008). In most cases, the *S. zooepidemicus* isolates examined contained at least one of these genetic loci: a predicted cell wall-anchored protein (SZO08560) with conserved domains found in internalins of *Listeria* species and/or a putative pilus cluster with similarities to the *rlrA* locus in *S. pneumoniae*. Expression of the SZO08560 gene was hypothesised to be switched 'on' and 'off' through a process of recombinase-mediated phase variation that has not yet been reported for a surface protein in streptococci. Switching between 2 phenotypes may aid the avoidance of adaptive immunity and is a hallmark of organisms persistently colonising mucosal surfaces. SZO08560 was significantly more prevalent in lower respiratory tract *S. zooepidemicus* isolates (including those isolated from cases of pneumonia, 100%, n = 14) than in upper respiratory tract isolates of *S. zooepidemicus* (including those associated with LN abscessation) (71%, n = 59) (P = 0.03, Fischer's exact test) and all other *S. zooepidemicus* isolates (69%, n = 124, P = 0.011, Fischer's exact test). Since this study, my colleagues at the AHT have examined more isolates of *S. zooepidemicus* and further demonstrated that a significantly higher proportion of isolates recovered from the lower respiratory tract or cases of pneumonia (93%, n = 41) had the SZO08560 gene compared to other *S. zooepidemicus* isolates (72%, n = 161, P = 0.004, Fischer's exact test, Andrew Waller, unpublished data). It is therefore proposed that a switch in phase is a virulence trait and the *in vitro* equine air-interface model described previously could be used to compare the ability of 'locked-on' and 'locked-off' mutants, either independently or during co-culture, to colonise and/or invade respiratory tissue (particularly lung tissue). Organ culture plugs from the air-interface model are usually homogenised to allow numbers of viable bacteria to be measured, but pre-treatment of plugs with antibiotics that target extracellular bacteria

could also be assessed in an attempt to quantify the number of intracellular bacteria. Phase variants are probably not ideal vaccine targets unless the 'off' phase coincides with reduced infection (Langermann & Ballou, 2001).

The 2nd pilus locus may also prove to be important to *S. zooepidemicus* infection, as a similar pilus contributes to the virulence of *S. pneumoniae* in mice following intranasal inoculation (Hava & Camilli, 2002). Properties conferred by this putative pilus could also be examined in the *in vitro* equine air-interface culture system and electron microscopy, using gold-conjugated antibodies recognising the backbone and ancillary proteins to analyse the structure of the pilus and the contributions of the different sortase enzymes, pilus protein components and their structural motifs to pilus assembly. Pilus proteins have shown promise as antigens for vaccine development (Buccato *et al.*, 2006; Gianfaldoni *et al.*, 2007; Maione *et al.*, 2005; Mora *et al.*, 2005; Telford *et al.*, 2006), although limited variation must exist in the target population if the vaccine is to provide broad coverage.

6.5 Has gene loss altered pilus structure and production in *S. equi*?

The same strategy could be employed to assay the 1st pilus locus in the *S. equi* subspecies, as the introduction of a stop mutation into a *tetR*-like gene is hypothesised to lead to a production of longer pili in *S. equi*. Basic transcriptional analysis could be used to establish the regulatory control exhibited by the TetR-like protein and electron microscopy used to compare pilus structure in wild-type *Se*4047, *Sz*H70 and various allelic replacement mutants containing either a repaired *tetR* regulator or deleted pilus subunits (in the parental *Se*4047 background). If this hypothesis is proven, then this pilus may play an important role in *S. equi* pathogenesis and its contribution to, for example, the rapid intracellular or paracellular translocation of *S. equi* across epithelial cells or biofilm formation in the GP, where *S. equi* has been shown to persist even in the absence of

chondroids, could be assessed. A pilus vaccine in *S. equi* may stand more chance of success than a similar vaccine in *S. zooepidemicus*, as the pilus target in former may prove to be less variable.

6.6 The influences of gene degradation and gene change on hyaluronate capsule production in *S. equi*

This study has highlighted other examples of gene degradation in *S. equi*, including the inactivation of the HA lyase gene, which is hypothesised to reduce the degradation of its own HA capsule and limit the spread of *S. equi* beyond the lymphatic system. Experiments could be performed to disrupt the same gene in *Sz*H70 and assess the impact this has on encapsulation and tissue spread using *in vitro* and *in vivo* assays. The HA gene in *Se*4047 and *S. zooepidemicus* ST-57 could also be repaired to see if these hyper-encapsulated strains gain a small colony phenotype like *Sz*H70 and reduced resistance to phagocytosis as a result.

The other main force that has shaped the genome evolution of *S. equi* is gene change. An intra-replichore inversion event has reorganised the genes of the *has* operon involved in HA biosynthesis. This change is hypothesised to alter the expression levels of precursors in HA biosynthesis in *Se*4047, which could also contribute to the hyper-encapsulated phenotype of *S. equi* and the significant resistance shown by this pathogen to phagocytosis. Quantitative RT-PCR analysis would reveal quickly any differences in the expression levels of these precursors between *Se*4047 and *Sz*H70.

6.7 Genetic variation in surface proteins: identification and characterisation of potential *S. equi*-specific vaccine targets

Many surface proteins varied in sequence between *Se*4047 and *Sz*H70/*Sz*MGCS10565, and were not screened by PCR in this study mostly because of the concern that variation in primer binding sites would mislead the interpretation of the results obtained. However, surface proteins are likely to be important in the interaction of these subspecies with their hosts and are good candidates for the development of subunit vaccines. Further sequencing of *S. zooepidemicus* strains is therefore required to assess the true extent of variation between *S. equi* and more closely-related strains of *S. zooepidemicus*. Antigens of *S. equi* that differ even when compared to their homologues in closely related *S. zooepidemicus* isolates may reflect the specific pathogenicity of *S. equi* and follow up work could aim to identify ligand binding activities and the contribution of individual antigens to adherence/invasion of host cells. It would be interesting to see if certain surface proteins contribute to colonisation in the GP, particularly as this site in the horse is so important to *S. equi* carriage and contains glycans that have not been reported in any sections of the equine respiratory tract (Parillo *et al.*, 2009). This study highlighted differences in potential integrin binding motifs in collagen-like surface proteins, which when mutated may reduce the ability of *S. equi* to bind and possibly invade certain cells.

6.8 The gain of ICE*Se2* by *S. equi*: a possible speciation event

Chapters 3 to 4 describe the recently published work characterising the ICE*Se2* NRPS in detail (Heather *et al.*, 2008). The gain of ICE*Se2* was identified as possibly the genetic event that defines the speciation of *S. equi*. ICE*Se2* encodes a novel siderophore-like NRPS system, the first of its kind to be identified in streptococci. In fact, NRPSs in general are rare in streptococci and other bacteria with genomes ≤3 Mb in size probably due to their large metabolic cost (Donadio *et al.*, 2007), which suggests that the *S. equi*

NRPS offers a critical function to this pathogen. As the secreted product of this NRPS system was shown to be involved in iron uptake, I hypothesise that iron acquisition by this process plays a vital role in *S. equi* pathogenesis. Although this was not demonstrated in a mouse infection model, its contribution to virulence would be best assessed in the natural host using an established strangles model (Hamilton *et al.*, 2006). The air interface organ culture could also be used to examine the influence of the NRPS on *S. equi* replication in an environment that is iron limiting.

The precise mechanism of ICE*Se2* NRPS product function is currently unclear as I have not yet identified direct iron-binding activity by this metabolite. Elucidating the structure of the NRPS product will be an important step towards understanding its activity. However, *S. equi*, during its persistence in LN abscesses and the GP, is likely to encounter lactoferrin, the iron-sequestering glycoprotein secreted by neutrophils. *S. pyogenes*, which is not able to produce its own siderophores, is also unable to use lactoferrin as an iron source (Eichenbaum *et al.*, 1996). It would therefore be interesting to see if the ICE*Se2* NRPS product allows *S. equi* to release iron from lactoferrin, particularly as the host catecholamine norepinephrine, although unable to remove iron from the CAS complex, can increase the availability of free iron to bacteria through an interaction that interferes with host glycoprotein iron sequestration (Freestone *et al.*, 2000). Norepinephrine formed stable complexes with lactoferrin and transferrin, and incubation of either of these glycoproteins with norepinephrine resulted in their loss of bound iron as indicated by the appearance of monoferric and apo-isoforms upon electrophoresis in denaturing gels (Freestone *et al.*, 2000). Similar experiments could be performed with the *eqb* NRPS product, using the purified peptide if this becomes available or conditioned media. Alternatively, the ability to incorporate ^{55}Fe from ^{55}Fe-labelled lactoferrin or transferrin could be measured in the different *S. equi* stains (Freestone *et al.*, 2000).

Genes required for the biosynthesis of the NRPS product were identified following its heterologous expression in *E. coli* and the deletion of the *eqbE* gene in *S. equi*. I proposed a mechanism for the biosynthesis of the NRPS product following structure-based bioinformatic analysis and hypothesised that the final equibactin product would contain a single residue of salicylate and 3 thiazoli(di)ne rings formed from the heterocyclisation of cysteine. Salicylate was proven to be a substrate for this biosynthetic process, but further work is now required to establish the contribution of cysteine (and any other substrates) and determine the precise mechanisms of biosynthesis, so that the structure of the product can finally be elucidated. *In vitro* reconstitution (Chapter 4) is likely to provide the best opportunity to reveal the structure of the NRPS product by HPLC, MS and NMR (nuclear magnetic resonance), given that previous attempts to identify the product in conditioned media from *S. equi* mutant strains or *E. coli* expressing the biosynthetic proteins have failed. Other researchers were able to reveal the complete biosynthetic pathway for yersiniabactin production using *in vitro* reconstitution (Miller *et al.*, 2002). Individual catalytic domains were omitted in order to define their function and deuterated substrates were used in order to validate the number of specific substrate molecules present in the final product. Similar analysis could be performed with the *S. equi* NRPS, in addition to ATP/[^{32}P]pyrophosphate exchange assays to assess A domain substrate specificity (Otten *et al.*, 2007). A method for the large scale purification of the NRPS product may be easier to develop once the molecular weight and/or structure of the molecule is defined. Biological assays could then be performed with purified NRPS product rather than conditioned media and X-ray crystallography could be used (with NMR data) to elucidate its precise metal binding chemistry.

Thin layer chromatography (TLC) could be used as an alternative approach to identifying the *eqb* NRPS product. Mutant *S. equi* strains that either overproduce (Δ*eqbHIJA*) or fail to make the metabolite (Δ*eqbAE*) could be grown in CDM supplemented with ^{14}C

salicylate, a known substrate of the NRPS. Following TLC and autoradiography, spots that are only present in the former strain could be scraped from the TLC plate, dissolved in a suitable solvent and assessed by MS. Rf values may need to be calculated and non-radioactive samples run separately or alongside to avoid contamination of MS equipment with radioactive material. The identity of a yersiniabactin spot from a TLC plate was previously assessed using atmospheric pressure chemical ionization MS (Ferreras *et al.*, 2005) and MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) MS (Miller *et al.*, 2002).

6.8.1 Mechanisms for transport of the *eqb* NRPS product(s)

ABC transporter(s) EqbK and/or EqbL are hypothesised to export the *eqb* NRPS product and a failure to export the product is hypothesised to result in excessive iron accumulation through the upregulation of alternative iron import mechanisms. These hypotheses would be best examined using separate Δ*eqbK*, Δ*eqbL*, Δ*eqbKA*, Δ*eqbLA* deletion strains in growth, streptonigrin and ^{55}Fe incorporation assays described in Chapter 3. This will establish whether or not *eqbK* and *eqbL* encode distinct homodimeric ABC transporters with distinct functions or a single heterodimeric ABC transporter with a single function. If a failure to export the NRPS product is proven to result in excessive ^{55}Fe accumulation, then global gene expression analysis could be used to investigate upregulated iron import mechanisms. Gene arrays have not yet been developed for *S. equi*/*S. zooepidemicus* but could be custom-made relatively easily using information from the genome analysis.

The ABC transporter (EqbH, I and/or J) was shown to be critical for the majority of NRPS product-mediated iron import and an as yet unidentified receptor is hypothesised to contribute to this process. *eqbH*, *I* and *J* are also hypothesised to form a single ABC transporter. The latter hypothesis can be assessed relatively easily using separate mutants

with deletions in these individual components as described above for *eqbK* and *eqbL*. However, examining the requirement for a separately encoded import receptor is likely to be more problematic. Pre- and post-infection sera collected from horses could be assessed for the ability of antibodies in the latter to neutralise import function. Blocking of NRPS product-mediated iron import would point to an extracellular import domain. A *S. equi* HIMAR1 transposon library (May *et al.*, 2004) generated in the Δ*eqbAE* mutant and grown on media supplemented with streptonigrin and Δ*eqbHIJA* conditioned media (positive selection) could be used to identify a cell surface receptor for the NRPS product.

The hypothesis could be disproven if an extracellular binding domain was shown to be encoded within *eqbHIJ* (for example in the C-terminal hydrophilic region of EqbI). A plasmid expressing *eqbHIJ* could be introduced into *Sz*H70 and other streptococcal/lactococcal bacterial species and their ability to acquire iron evaluated following cross-feeding with conditioned media from the *S. equi* Δ*eqbHIJA* strain (compared to media from the Δ*eqbAE* strain). If these recombinant strains all showed increased *eqb* NRPS product-mediated iron accumulation, then either they all share a suitable receptor or, more likely, an extracytoplasmic domain within EqbHIJ scavenges the NRPS product for import. Biochemical, immunological and genetic procedures could be used to determine whether or not the hydrophilic C-terminal region of EqbI is exposed at the extracellular surface. For example, the fusion of a protein tag or β-lactamase into this region of EqbI could be used to screen for extracellular exposure using immunofluoresence or ampicillin resistance (Pearce *et al.*, 1992), respectively. Alternatively, import might occur without the involvement of a solute binding domain. However, no change in streptonigrin sensitivity or ^{55}Fe incorporation in some or all of the cross-fed recombinant species expressing EqbHIJ would suggest that a receptor is required for import of the NRPS product and that this is encoded in a *S. equi*-specific genomic region relative to the recombinant strains showing no change in iron accumulation.

6.8.2 A novel IdeR-like iron dependent repressor

The *eqbE* biosynthetic gene of the NRPS locus was repressed by the actions of a repressor (EqbA) shown to be responsive to Fe^{2+}, Zn^{2+} and to a lesser extent Mn^{2+}. This novel repressor is hypothesised to represent a new variant of the IdeR global iron-dependent repressor superfamily in Gram positive bacteria. Attempts could be made to crystallise EqbA and elucidate the structure of this novel *S. equi* IdeR-like iron-dependent repressor. This would require purification of recombinant EqbA in a suitable buffer to a scale and purity required to yield crystals that diffract X-rays to sufficient resolution to allow structure determination. Genetic analysis could then be performed to validate the contribution of specific residues to functions such as operator binding, dimerisation, other conformational changes and metal binding/selectivity. DNA footprinting could be used to define the exact region of the eqbB promoter bound by EqbA.

eqbB-N is hypothesised to be transcribed as a single operon (polycistronic mRNA molecule) under the repression of EqbA in the presence of iron. Further quantitative RT-PCR experiments could be performed assess this hypothesis using NTA chelation to remove iron from growth media. The EqbA repressor is also hypothesised regulate gene(s) outside of the *eqb* cluster as slight differences in growth rate were seen between the Δ*eqbAE* and Δ*eqbE S. equi* strains, which appeared to be independent of intracellular iron concentrations (suggested by streptonigrin sensitivity). Any regulatory control exhibited by EqbA outside of the *eqb* NRPS could be investigated following culture *in vitro* and comparing global gene expression profiles in the Δ*eqbAE* and Δ*eqbE S. equi* strains. The general effects of *eqb* NRPS product production and disregulation of iron import mechanisms of *S. equi* could also be determined using gene arrays and various gene deletion/wildtype strains.

6.8.3 Is ICE*Se2* capable of mobilisation?

S. equi is proposed to have gained the *eqb* NRPS though conjugation. Because circular intermediates of ICE*Se2* were not detected by PCR in this study, I hypothesise that this *S. equi* element no longer retains the ability to recircularise and transfer to other bacteria. Various stimuli including DNA-damaging treatments used for the activation of the phage lytic pathway and other ICEs could be assessed for their ability to induce conjugative transfer of ICE*Se2* and disprove this hypothesis (Auchtung *et al.*, 2007; Beaber *et al.*, 2004; Bellanger *et al.*, 2009; Bonheyo *et al.*, 2001). Mating experiments could also be performed to see if ICE*Se2* and an antibiotic resistance cassette incorporated into this element can be transferred from *Se*4047 into *S. zooepidemicus* or other bacteria (Bellanger *et al.*, 2009).

6.8.4 The development of *S. equi* diagnostic tools and vaccines specific to ICE*Se2*

Given its absence from the *S. zooepidemicus* population, the *eqbE* gene of the ICE*Se2* NRPS was used as a target for the successful development of sensitive, specific PCR and QPCR tests for the detection of *S. equi* (Chapter 5). This QPCR is currently being used commercially and has received positive feedback, particularly in terms of turnaround time.

In terms of vaccine development, either a receptor for the NRPS product (if identified) could be investigated for its efficacy as a subunit vaccine or deletions in *eqb* genes could be assessed for their contribution to live attenuated vaccine strains. The NRPS product itself is hypothesised to be too small to generate an antibody response; this could be tested by comparing the capacity of pre- and post-infection sera collected from horses to neutralise NRPS product function in streptonigrin and ^{55}Fe accumulation assays.

6.9 Concluding remarks

Overall, this study has significantly increased our understanding of *S. equi* evolution and the complex interplay of genetic forces that have shaped the transition of this pathogen from coloniser to invader. A wealth of information is now available that can be utilised to examine in detail the many potential virulence factors and aspects of *S. equi* pathogenicity. A number of possible vaccine targets have been identified that could be investigated further for their suitability to the development of live attenuated or subunit vaccines to provide broad spectrum, efficacious protection against strangles. Our understanding of *S. zooepidemicus* has also been widened to reveal many features that help to explain the versatile nature of this subspecies. The sheer diversity of the *S. zooepidemicus* population is likely to hinder the development of suitable vaccines against important diseases associated with this opportunistic pathogen. However, some avenues of research have already been opened and the further analysis of many more genomes may help to identify invariant targets that contribute to pathogenicity in subsets of the *S. zooepidemicus* population.

References

Abbot, E. L., Smith, W. D., Siou, G. P., Chiriboga, C., Smith, R. J., Wilson, J. A., Hirst, B. H. & Kehoe, M. A. (2007). Pili mediate specific adhesion of Streptococcus pyogenes to human tonsil and skin. *Cell Microbiol* **9**, 1822-1833.

Abraham, J. M., Freitag, C. S., Clements, J. R. & Eisenstein, B. I. (1985). An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc Natl Acad Sci U S A* **82**, 5724-5727.

Ajdic, D., McShan, W. M., McLaughlin, R. E. & other authors (2002). Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen. *Proc Natl Acad Sci U S A* **99**, 14434-14439.

Al-Ghamdi, G. M., Kapur, V., Ames, T. R., Timoney, J. F., Love, D. N. & Mellencamp, M. A. (2000). Use of repetitive sequence-based polymerase chain reaction for molecular epidemiologic analysis of Streptococcus equi subspecies equi. *Am J Vet Res* **61**, 699-705.

Alber, J., El-Sayed, A., Lammler, C., Hassan, A. A., Weiss, R. & Zschock, M. (2004). Multiplex polymerase chain reaction for identification and differentiation of Streptococcus equi subsp. zooepidemicus and Streptococcus equi subsp. equi. *J Vet Med B Infect Dis Vet Public Health* **51**, 455-458.

Alber, J., El-Sayed, A., Estoepangestie, S., Lammler, C. & Zschock, M. (2005). Dissemination of the superantigen encoding genes seeL, seeM, szeL and szeM in Streptococcus equi subsp. equi and Streptococcus equi subsp. zooepidemicus. *Vet Microbiol* **109**, 135-141.

Alice, A. F., Lopez, C. S. & Crosa, J. H. (2005). Plasmid- and chromosome-encoded redundant and specific functions are involved in biosynthesis of the siderophore anguibactin in Vibrio anguillarum 775: a case of chance and necessity? *J Bacteriol* **187**, 2209-2214.

Allard, K. A., Viswanathan, V. K. & Cianciotto, N. P. (2006). lbtA and lbtB are required for production of the Legionella pneumophila siderophore legiobactin. *J Bacteriol* **188**, 1351-1363.

Alm, R. A., Ling, L. S., Moir, D. T. & other authors (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. *Nature* **397**, 176-180.

Andersson, S. G., Zomorodipour, A., Andersson, J. O. & other authors (1998). The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* **396**, 133-140.

Anisimov, R., Brem, D., Heesemann, J. & Rakin, A. (2005a). Molecular mechanism of YbtA-mediated transcriptional regulation of divergent overlapping promoters ybtA and irp6 of Yersinia enterocolitica. *FEMS Microbiol Lett* **250**, 27-32.

Anisimov, R., Brem, D., Heesemann, J. & Rakin, A. (2005b). Transcriptional regulation of high pathogenicity island iron uptake genes by YbtA. *Int J Med Microbiol* **295**, 19-28.

Anzai, T., Sheoran, A. S., Kuwamoto, Y., Kondo, T., Wada, R., Inoue, T. & Timoney, J. F. (1999a). Streptococcus equi but not Streptococcus zooepidemicus produces potent mitogenic responses from equine peripheral blood mononuclear cells. *Vet Immunol Immunopathol* **67**, 235-246.

Anzai, T., Timoney, J. F., Kuwamoto, Y., Fujita, Y., Wada, R. & Inoue, T. (1999b). In vivo pathogenicity and resistance to phagocytosis of Streptococcus equi strains with different levels of capsule expression. *Vet Microbiol* **67**, 277-286.

Anzai, T., Kuwamoto, Y., Wada, R., Sugita, S., Kakuda, T., Takai, S., Higuchi, T. & Timoney, J. F. (2005). Variation in the N-terminal region of an M-like protein of Streptococcus equi and evaluation of its potential as a tool in epidemiologic studies. *Am J Vet Res* **66**, 2167-2171.

Aras, R. A., Kang, J., Tschumi, A. I., Harasaki, Y. & Blaser, M. J. (2003). Extensive repetitive DNA facilitates prokaryotic genome plasticity. *Proc Natl Acad Sci U S A* **100**, 13579-13584.

Arcus, V. L., Proft, T., Sigrell, J. A., Baker, H. M., Fraser, J. D. & Baker, E. N. (2000). Conservation and variation in superantigen structure and activity highlighted by the three-dimensional structures of two new superantigens from Streptococcus pyogenes. *J Mol Biol* **299**, 157-168.

Artiushin, S. C., Timoney, J. F., Sheoran, A. S. & Muthupalani, S. K. (2002). Characterization and immunogenicity of pyrogenic mitogens SePE-H and SePE-I of *Streptococcus equi*. *Microb Pathog* **32**, 71-85.

Asbury, A. C., Gorman, N. T. & Foster, G. W. (1984). Uterine defense mechanisms in the mare: Serum opsonins affecting phagocytosis of Streptococcus zooepidemicus by equine neutrophils. *Theriogenology* **21**, 375-385.

Auchtung, J. M., Lee, C. A., Garrison, K. L. & Grossman, A. D. (2007). Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICEBs1 of Bacillus subtilis. *Mol Microbiol* **64**, 1515-1528.

Aziz, R. K., Edwards, R. A., Taylor, W. W., Low, D. E., McGeer, A. & Kotb, M. (2005). Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1T1 clone of Streptococcus pyogenes. *J Bacteriol* **187**, 3311-3318.

Bailey, C. J., Reid, S. W., Hodgson, D. R. & Rose, R. J. (1999). Impact of injuries and disease on a cohort of two- and three-year-old thoroughbreds in training. *Vet Rec* **145**, 487-493.

Baker, J. R., Dong, S. & Pritchard, D. G. (2002). The hyaluronan lyase of *Streptococcus pyogenes* bacteriophage H4489A. *Biochem J* **365**, 317-322.

Balter, S., Benin, A., Pinto, S. W. & other authors (2000). Epidemic nephritis in Nova Serrana, Brazil. *Lancet* **355**, 1776-1780.

Bankhead, T. & Chaconas, G. (2007). The role of VlsE antigenic variation in the Lyme disease spirochete: persistence through a mechanism that differs from other pathogens. *Mol Microbiol* **65**, 1547-1558.

Banks, D. J., Lei, B. & Musser, J. M. (2003). Prophage induction and expression of prophage-encoded virulence factors in group A *Streptococcus* serotype M3 strain MGAS315. *Infect Immun* **71**, 7079-7086.

Banks, D. J., Porcella, S. F., Barbian, K. D. & other authors (2004). Progress toward characterization of the group A Streptococcus metagenome: complete genome sequence of a macrolide-resistant serotype M6 strain. *J Infect Dis* **190**, 727-738.

Bannister, M. F., Benson, C. E. & Sweeney, C. R. (1985). Rapid species identification of group C streptococci isolated from horses. *J Clin Microbiol* **21**, 524-526.

Barchini, E. & Cowart, R. E. (1996). Extracellular iron reductase activity produced by Listeria monocytogenes. *Arch Microbiol* **166**, 51-57.

Barnett, T. C., Patel, A. R. & Scott, J. R. (2004). A novel sortase, SrtC2, from Streptococcus pyogenes anchors a surface protein containing a QVPTGV motif to the cell wall. *J Bacteriol* **186**, 5865-5875.

Barnham, M., Cole, G., Efstratiou, A., Tagg, J. R. & Skjold, S. A. (1987). Characterization of Streptococcus zooepidemicus (Lancefield group C) from human and selected animal infections. *Epidemiol Infect* **98**, 171-182.

Barocchi, M. A., Ries, J., Zogaj, X. & other authors (2006). A pneumococcal pilus influences virulence and host inflammatory responses. *Proc Natl Acad Sci U S A* **103**, 2857-2862.

Barquero, N., Chanter, N., Laxton, R., Wood, J. L. & Richard Newton, J. (2009). Molecular epidemiology of Streptococcus zooepidemicus isolated from the respiratory tracts of Thoroughbred racehorses in training. *Vet J.*

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712.

Bartelt, M. A. & Duncan, J. L. (1978). Adherence of group A streptococci to human epithelial cells. *Infect Immun* **20**, 200-208.

Bateman, A., Birney, E., Cerruti, L. & other authors (2002). The Pfam protein families database. *Nucleic Acids Res* **30**, 276-280.

Baverud, V., Johansson, S. K. & Aspan, A. (2007). Real-time PCR for detection and differentiation of Streptococcus equi subsp. equi and Streptococcus equi subsp. zooepidemicus. *Vet Microbiol* **124**, 219-229.

Bazely, P. L. (1942). Studies with equine streptococci: 4. Cross-immunity to S. equi. *Aust Vet J* **18**, 189-194.

Beaber, J. W., Hochhut, B. & Waldor, M. K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**, 72-74.

Bearden, S. W., Fetherston, J. D. & Perry, R. D. (1997). Genetic organization of the yersiniabactin biosynthetic region and construction of avirulent mutants in *Yersinia pestis*. *Infect Immun* **65**, 1659-1668.

Beare, P. A., For, R. J., Martin, L. W. & Lamont, I. L. (2003). Siderophore-mediated cell signalling in Pseudomonas aeruginosa: divergent pathways regulate virulence factor production and siderophore receptor synthesis. *Mol Microbiol* **47**, 195-207.

Bellanger, X., Roberts, A. P., Morel, C., Choulet, F., Pavlovic, G., Mullany, P., Decaris, B. & Guedon, G. (2009). Conjugative transfer of the integrative conjugative elements ICESt1 and ICESt3 from Streptococcus thermophilus. *J Bacteriol.*

Beres, S. B., Sylva, G. L., Barbian, K. D. & other authors (2002). Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc Natl Acad Sci U S A* **99**, 10078-10083.

Beres, S. B., Richter, E. W., Nagiec, M. J., Sumby, P., Porcella, S. F., DeLeo, F. R. & Musser, J. M. (2006). Molecular genetic anatomy of inter- and intraserotype variation in the human bacterial pathogen group A *Streptococcus*. *Proc Natl Acad Sci U S A* **103**, 7059-7064.

Beres, S. B. & Musser, J. M. (2007). Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. *PLoS ONE* **2**, e800.

Beres, S. B., Sesso, R., Pinto, S. W., Hoe, N. P., Porcella, S. F., Deleo, F. R. & Musser, J. M. (2008). Genome sequence of a lancefield group C *Streptococcus zooepidemicus* strain causing epidemic nephritis: new information about an old disease. *PLoS ONE* **3**, e3026.

Berish, S. A., Subbarao, S., Chen, C. Y., Trees, D. L. & Morse, S. A. (1993). Identification and cloning of a fur homolog from Neisseria gonorrhoeae. *Infect Immun* **61**, 4599-4606.

Bessen, D. E. & Kalia, A. (2002). Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in Streptococcus pyogenes. *Infect Immun* **70**, 1159-1167.

Blank, L. M., Hugenholtz, P. & Nielsen, L. K. (2008). Evolution of the hyaluronic acid synthesis (*has*) operon in *Streptococcus zooepidemicus* and other pathogenic streptococci. *J Mol Evol.*

Bobrov, A. G., Geoffroy, V. A. & Perry, R. D. (2002). Yersiniabactin production requires the thioesterase domain of HMWP2 and YbtD, a putative phosphopantetheinylate transferase. *Infect Immun* **70**, 4204-4214.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D. & Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403. *Genome Res* **11**, 731-753.

Bolotin, A., Quinquis, B., Renault, P. & other authors (2004). Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. *Nat Biotechnol* **22**, 1554-1558.

Bonheyo, G. T., Hund, B. D., Shoemaker, N. B. & Salyers, A. A. (2001). Transfer region of a Bacteroides conjugative transposon contains regulatory as well as structural genes. *Plasmid* **46**, 202-209.

Bordes-Benitez, A., Sanchez-Onoro, M., Suarez-Bordon, P., Garcia-Rojas, A. J., Saez-Nieto, J. A., Gonzalez-Garcia, A., Alamo-Antunez, I., Sanchez-Maroto, A. & Bolanos-Rivero, M. (2006). Outbreak of Streptococcus equi subsp. zooepidemicus infections on the island of Gran Canaria associated with the consumption of inadequately pasteurized cheese. *Eur J Clin Microbiol Infect Dis* **25**, 242-246.

Bormann, N. E. & Cleary, P. P. (1997). Transcriptional analysis of mga, a regulatory gene in Streptococcus pyogenes: identification of monocistronic and bicistronic transcripts that phase vary. *Gene* **200**, 125-134.

Boschwitz, J. S. & Timoney, J. F. (1994a). Inhibition of C3 deposition on Streptococcus equi subsp. equi by M protein: a mechanism for survival in equine blood. *Infect Immun* **62**, 3515-3520.

Boschwitz, J. S. & Timoney, J. F. (1994b). Characterization of the antiphagocytic activity of equine fibrinogen for Streptococcus equi subsp. equi. *Microb Pathog* **17**, 121-129.

Bossi, L., Fuentes, J. A., Mora, G. & Figueroa-Bossi, N. (2003). Prophage contribution to bacterial population dynamics. *J Bacteriol* **185**, 6467-6471.

Bradley, S. F., Gordon, J. J., Baumgartner, D. D., Marasco, W. A. & Kauffman, C. A. (1991). Group C streptococcal bacteremia: analysis of 88 cases. *Rev Infect Dis* **13**, 270-280.

Brickman, T. J. & Armstrong, S. K. (2005). Bordetella AlcS transporter functions in alcaligin siderophore export and is central to inducer sensing in positive regulation of alcaligin system gene expression. *J Bacteriol* **187**, 3650-3661.

Brooks, D. E., Andrew, S. E., Biros, D. J., Denis, H. M., Cutler, T. J., Strubbe, D. T. & Gelatt, K. N. (2000). Ulcerative keratitis caused by beta-hemolytic Streptococcus equi in 11 horses. *Vet Ophthalmol* **3**, 121-125.

Brouillard, J. N., Gunther, S., Varma, A. K. & other authors (2007). Crystal structure of the streptococcal superantigen SpeI and functional role of a novel loop domain in T cell activation by group V superantigens. *J Mol Biol* **367**, 925-934.

Brown, J. S., Gilliland, S. M., Ruiz-Albert, J. & Holden, D. W. (2002). Characterization of pit, a Streptococcus pneumoniae iron uptake ABC transporter. *Infect Immun* **70**, 4389-4398.

Brown, J. S. & Holden, D. W. (2002). Iron acquisition by Gram-positive bacterial pathogens. *Microbes Infect* **4**, 1149-1156.

Brussow, H., Canchaya, C. & Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**, 560-602.

Buccato, S., Maione, D., Rinaudo, C. D., Volpini, G., Taddei, A. R., Rosini, R., Telford, J. L., Grandi, G. & Margarit, I. (2006). Use of Lactococcus lactis expressing pili from group B Streptococcus as a broad-coverage vaccine against streptococcal disease. *J Infect Dis* **194**, 331-340.

Buchanan, S. K., Smith, B. S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D. & Deisenhofer, J. (1999). Crystal structure of the outer membrane active transporter FepA from Escherichia coli. *Nat Struct Biol* **6**, 56-63.

Burrell, M. H., Wood, J. L., Whitwell, K. E., Chanter, N., Mackintosh, M. E. & Mumford, J. A. (1996). Respiratory disease in thoroughbred horses in training: the

relationships between disease and viruses, bacteria and environment. *Vet Rec* **139**, 308-313.

Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. (2002a). The ICESt1 element of Streptococcus thermophilus belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* **48**, 77-97.

Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. (2002b). Conjugative transposons: the tip of the iceberg. *Mol Microbiol* **46**, 601-610.

Burrus, V. & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol* **155**, 376-386.

Burts, M. L., Williams, W. A., DeBord, K. & Missiakas, D. M. (2005). EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc Natl Acad Sci U S A* **102**, 1169-1174.

Burts, M. L., DeDent, A. C. & Missiakas, D. M. (2008). EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of Staphylococcus aureus. *Mol Microbiol* **69**, 736-746.

Butler, A. R., Bate, N. & Cundliffe, E. (1999). Impact of thioesterase activity on tylosin biosynthesis in Streptomyces fradiae. *Chem Biol* **6**, 287-292.

Camilli, R., Del Grosso, M., Iannelli, F. & Pantosti, A. (2008). New genetic element carrying the erythromycin resistance determinant erm(TR) in Streptococcus pneumoniae. *Antimicrob Agents Chemother* **52**, 619-625.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. (2003). Prophage genomics. *Microbiol Mol Biol Rev* **67**, 238-276, table of contents.

Cardwell, J. M. (2007). An epidemiological study of inflammatory airway disease in national hunt racehorses in the United Kingdom. In *Life Sciences*: Open University.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422-3423.

Caswell, C. C., Lukomska, E., Seo, N. S., Hook, M. & Lukomski, S. (2007). Scl1-dependent internalization of group A Streptococcus via direct interactions with the alpha2beta(1) integrin enhances pathogen survival and re-emergence. *Mol Microbiol* **64**, 1319-1331.

Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E. & Lukomski, S. (2008a). Identification of the First Prokaryotic Collagen Sequence Motif That Mediates Binding to Human Collagen Receptors, Integrins {alpha}2{beta}1 and {alpha}11{beta}1. *J Biol Chem* **283**, 36168-36175.

Caswell, C. C., Han, R., Hovis, K. M., Ciborowski, P., Keene, D. R., Marconi, R. T. & Lukomski, S. (2008b). The Scl1 protein of M6-type group A Streptococcus binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement. *Mol Microbiol* **67**, 584-596.

Causey, R. C., Paccamonti, D. L. & Todd, W. J. (1995). Antiphagocytic properties of uterine isolates of Streptococcus zooepidemicus and mechanisms of killing in freshly obtained blood of horses. *Am J Vet Res* **56**, 321-328.

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C. & other authors (2005). Extensive DNA inversions in the B. fragilis genome control variable gene expression. *Science* **307**, 1463-1465.

Chalker, V. J., Brooks, H. W. & Brownlie, J. (2003). The association of Streptococcus equi subsp. zooepidemicus with canine infectious respiratory disease. *Vet Microbiol* **95**, 149-156.

Challis, G. L. & Ravel, J. (2000). Coelichelin, a new peptide siderophore encoded by the Streptomyces coelicolor genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiol Lett* **187**, 111-114.

Challis, G. L., Ravel, J. & Townsend, C. A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**, 211-224.

Challis, G. L. & Naismith, J. H. (2004). Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* **14**, 748-756.

Challis, G. L. (2006). Engineering Escherichia coli to produce nonribosomal peptide antibiotics. *Nat Chem Biol* **2**, 398-400.

Chang, G. & Roth, C. B. (2001). Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**, 1793-1800.

Chang, G. (2003). Structure of MsbA from Vibrio cholera: a multidrug resistance ABC transporter homolog in a closed conformation. *J Mol Biol* **330**, 419-430.

Chanter, N., Smith, K. C. & Mumford, J. A. (1995). Equine strangles modelled in mice. *Vet Microbiol* **43**, 209-218.

Chanter, N. (1997). Streptococci and enterococci as animal pathogens. *Soc Appl Bacteriol Symp Ser* **26**, 100S-109S.

Chanter, N., Collin, N., Holmes, N., Binns, M. & Mumford, J. (1997). Characterization of the Lancefield group C streptococcus 16S-23S RNA gene intergenic spacer and its potential for identification and sub-specific typing. *Epidemiol Infect* **118**, 125-135.

Chanter, N., Ward, C. L., Talbot, N. C., Flanagan, J. A., Binns, M., Houghton, S. B., Smith, K. C. & Mumford, J. A. (1999). Recombinant hyaluronate associated protein as a protective immunogen against Streptococcus equi and Streptococcus zooepidemicus challenge in mice. *Microb Pathog* **27**, 133-143.

Chanter, N., Talbot, N. C., Newton, J. R., Hewson, D. & Verheyen, K. (2000). Streptococcus equi with truncated M-proteins isolated from outwardly healthy horses. *Microbiology* **146 (Pt 6)**, 1361-1369.

Chanter, N., Collin, N. C. & Mumford, J. A. (1994). Equine Infectious Diseases VII. In *Proceedings of the Seventh International Conference, Tokyo, Japan, 8th-11th June 1994*, pp. 201-210. Edited by H. P. Nakajima, W. Newmarket: R & W Publications.

Chapman, P. S., Green, C., Main, J. P., Taylor, P. M., Cunningham, F. M., Cook, A. J. & Marr, C. M. (2000). Retrospective study of the relationships between age, inflammation and the isolation of bacteria from the lower respiratory tract of thoroughbred horses. *Vet Rec* **146**, 91-95.

Chenault, S. S. & Earhart, C. F. (1991). Organization of genes encoding membrane proteins of the Escherichia coli ferrienterobactin permease. *Mol Microbiol* **5**, 1405-1413.

Chhatwal, G. S., Dutra, I. S. & Blobel, H. (1985). Fibrinogen binding inhibits the fixation of the third component of human complement on surface of groups A, B, C, and G streptococci. *Microbiol Immunol* **29**, 973-980.

Chmouryguina, I., Suvorov, A., Ferrieri, P. & Cleary, P. P. (1996). Conservation of the C5a peptidase genes in group A and B streptococci. *Infect Immun* **64**, 2387-2390.

Chou, C. J., Wisedchaisri, G., Monfeli, R. R., Oram, D. M., Holmes, R. K., Hol, W. G. & Beeson, C. (2004). Functional studies of the Mycobacterium tuberculosis iron-dependent regulator. *J Biol Chem* **279**, 53554-53561.

Christley, R. M., Hodgson, D. R., Rose, R. J., Hodgson, J. L., Wood, J. L. & Reid, S. W. (2001). Coughing in thoroughbred racehorses: risk factors and tracheal endoscopic and cytological findings. *Vet Rec* **148**, 99-104.

Clancy, A., Loar, J. W., Speziali, C. D., Oberg, M., Heinrichs, D. E. & Rubens, C. E. (2006). Evidence for siderophore-dependent iron acquisition in group B streptococcus. *Mol Microbiol* **59**, 707-721.

Cleary, P. P., Prahbu, U., Dale, J. B., Wexler, D. E. & Handley, J. (1992). Streptococcal C5a peptidase is a highly specific endopeptidase. *Infect Immun* **60**, 5219-5223.

Cobessi, D., Celia, H. & Pattus, F. (2005). Crystal structure at high resolution of ferric-pyochelin and its membrane receptor FptA from Pseudomonas aeruginosa. *J Mol Biol* **352**, 893-904.

Cole, S. T., Eiglmeier, K., Parkhill, J. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007-1011.

Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**, 4174-4183.

Cowan, C., Jones, H. A., Kaya, Y. H., Perry, R. D. & Straley, S. C. (2000). Invasion of epithelial cells by Yersinia pestis: evidence for a Y. pestis-specific invasin. *Infect Immun* **68**, 4523-4530.

Crosa, J. H. (1997). Signal transduction and transcriptional and posttranscriptional control of iron-regulated genes in bacteria. *Microbiol Mol Biol Rev* **61**, 319-336.

Crosa, J. H. & Walsh, C. T. (2002). Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* **66**, 223-249.

Davidson, A. L. & Chen, J. (2004). ATP-binding cassette transporters in bacteria. *Annu Rev Biochem* **73**, 241-268.

Dawson, R. J. & Locher, K. P. (2006). Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180-185.

De Crecy-Lagard, V., Marliere, P. & Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *C R Acad Sci III* **318**, 927-936.

de Crecy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S., Famechon, A., Bamas-Jacques, N., Crouzet, J. & Thibaut, D. (1997). Pristinamycin I biosynthesis in Streptomyces pristinaespiralis: molecular characterization of the first two structural peptide synthetase genes. *J Bacteriol* **179**, 705-713.

Deng, W., Burland, V., Plunkett, G., 3rd & other authors (2002). Genome sequence of Yersinia pestis KIM. *J Bacteriol* **184**, 4601-4611.

DiGiuseppe Champion, P. A. & Cox, J. S. (2007). Protein secretion systems in Mycobacteria. *Cell Microbiol* **9**, 1376-1384.

Divers, T. J., Timoney, J. F., Lewis, R. M. & Smith, C. A. (1992). Equine glomerulonephritis and renal failure associated with complexes of group-C streptococcal antigen and IgG antibody. *Vet Immunol Immunopathol* **32**, 93-102.

Dixon, S., Haswell, M., Harrington, D. & Sutcliffe, I. C. (2001). Surface immunolocalisation of HPr in the equine pathogen Streptococcus equi. *Syst Appl Microbiol* **24**, 486-489.

Donadio, S. & Sosio, M. (2003). Strategies for combinatorial biosynthesis with modular polyketide synthases. *Comb Chem High Throughput Screen* **6**, 489-500.

Donadio, S., Monciardini, P. & Sosio, M. (2007). Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* **24**, 1073-1109.

Downar, J., Willey, B. M., Sutherland, J. W., Mathew, K. & Low, D. E. (2001). Streptococcal meningitis resulting from contact with an infected horse. *J Clin Microbiol* **39**, 2358-2359.

Dramsi, S., Trieu-Cuot, P. & Bierne, H. (2005). Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. *Res Microbiol* **156**, 289-297.

Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prevost, M. C., Kojadinovic, M., Lalioui, L., Poyart, C. & Trieu-Cuot, P. (2006). Assembly and role of pili in group B streptococci. *Mol Microbiol* **60**, 1401-1413.

Du, L., Sanchez, C., Chen, M., Edwards, D. J. & Shen, B. (2000). The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003

supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem Biol* **7**, 623-642.

Duerfahrt, T., Eppelmann, K., Muller, R. & Marahiel, M. A. (2004). Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. *Chem Biol* **11**, 261-271.

Edwards, R. J., Taylor, G. W., Ferguson, M. & other authors (2005). Specific C-terminal cleavage and inactivation of interleukin-8 by invasive disease isolates of Streptococcus pyogenes. *J Infect Dis* **192**, 783-790.

Eichenbaum, Z., Muller, E., Morse, S. A. & Scott, J. R. (1996). Acquisition of iron from host proteins by the group A streptococcus. *Infect Immun* **64**, 5428-5429.

Eisen, J. A., Heidelberg, J. F., White, O. & Salzberg, S. L. (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol* **1**, RESEARCH0011.

Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953-971.

Evans, A. C. (1936). Studies on Hemolytic Streptococci: III. Streptococcus equi and Related Strains. *J Bacteriol* **32**, 541-556.

Evans, S. L., Arceneaux, J. E., Byers, B. R., Martin, M. E. & Aranha, H. (1986). Ferrous iron transport in Streptococcus mutans. *J Bacteriol* **168**, 1096-1099.

Evers, W. D. (1968). Effect of furaltadone on strangles in horses. *J Am Vet Med Assoc* **152**, 1394-1398.

Facklam, R. (2002). What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin Microbiol Rev* **15**, 613-630.

Falker, S., Nelson, A. L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S. & Henriques-Normark, B. (2008). Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol Microbiol* **70**, 595-607.

Fan, H., Wang, Y., Tang, F. & Lu, C. (2008). Determination of the mimic epitope of the M-like protein adhesin in swine Streptococcus equi subsp. zooepidemicus. *BMC Microbiol* **8**, 170.

Farhana, A., Kumar, S., Rathore, S. S., Ghosh, P. C., Ehtesham, N. Z., Tyagi, A. K. & Hasnain, S. E. (2008). Mechanistic insights into a novel exporter-importer system of Mycobacterium tuberculosis unravel its role in trafficking of iron. *PLoS ONE* **3**, e2087.

Farrow, J. A. & Collins, M. D. (1984). DNA base composition, DNA-DNA homology and long-chain fatty acid studies on streptococcus thermophilus and Streptococcus salivarius. *J Gen Microbiol* **130**, 357-362.

Feng, Y., Chen, C. J., Su, L. H., Hu, S., Yu, J. & Chiu, C. H. (2008). Evolution and pathogenesis of Staphylococcus aureus: lessons learned from genotyping and comparative genomics. *FEMS Microbiol Rev* **32**, 23-37.

Feng, Z., Qi, J., Tsuge, T., Oba, Y., Kobayashi, T., Suzuki, Y., Sakagami, Y. & Ojika, M. (2005). Construction of a bacterial artificial chromosome library for a myxobacterium of the genus Cystobacter and characterization of an antibiotic biosynthetic gene cluster. *Biosci Biotechnol Biochem* **69**, 1372-1380.

Fernandez, E., Blume, V., Garrido, P., Collins, M. D., Mateos, A., Dominguez, L. & Fernandez-Garayzabal, J. F. (2004). Streptococcus equi subsp. ruminatorum subsp. nov., isolated from mastitis in small ruminants. *Int J Syst Evol Microbiol* **54**, 2291-2296.

Ferreras, J. A., Ryu, J. S., Di Lello, F., Tan, D. S. & Quadri, L. E. (2005). Small-molecule inhibition of siderophore biosynthesis in Mycobacterium tuberculosis and Yersinia pestis. *Nat Chem Biol* **1**, 29-32.

Fetherston, J. D., Bearden, S. W. & Perry, R. D. (1996). YbtA, an AraC-type regulator of the Yersinia pestis pesticin/yersiniabactin receptor. *Mol Microbiol* **22**, 315-325.

Fetherston, J. D., Bertolino, V. J. & Perry, R. D. (1999). YbtP and YbtQ: two ABC transporters required for iron uptake in Yersinia pestis. *Mol Microbiol* **32**, 289-299.

Fintl, C., Dixon, P. M., Brazil, T. J., Pirie, R. S. & McGorum, B. C. (2000). Endoscopic and bacteriological findings in a chronic outbreak of strangles. *Vet Rec* **147**, 480-484.

Fisher, M., Huang, Y. S., Li, X., McIver, K. S., Toukoki, C. & Eichenbaum, Z. (2008). Shr is a broad-spectrum surface receptor that contributes to adherence and virulence in group A streptococcus. *Infect Immun* **76**, 5006-5015.

Flanagan, J., Collin, N., Timoney, J., Mitchell, T., Mumford, J. A. & Chanter, N. (1998). Characterization of the haemolytic activity of Streptococcus equi. *Microb Pathog* **24**, 211-221.

Flannagan, S. E., Zitzow, L. A., Su, Y. A. & Clewell, D. B. (1994). Nucleotide sequence of the 18-kb conjugative transposon Tn916 from Enterococcus faecalis. *Plasmid* **32**, 350-354.

Flock, M., Jacobsson, K., Frykberg, L., Hirst, T. R., Franklin, A., Guss, B. & Flock, J. I. (2004). Recombinant Streptococcus equi proteins protect mice in challenge experiments and induce immune response in horses. *Infect Immun* **72**, 3228-3236.

Flock, M., Karlstrom, A., Lannergard, J., Guss, B. & Flock, J. I. (2006). Protective effect of vaccination with recombinant proteins from Streptococcus equi subspecies equi in a strangles model in the mouse. *Vaccine* **24**, 4144-4151.

Francis, A. J., Nimmo, G. R., Efstratiou, A., Galanis, V. & Nuttall, N. (1993). Investigation of milk-borne Streptococcus zooepidemicus infection associated with glomerulonephritis in Australia. *J Infect* **27**, 317-323.

Franco, A. A. (2004). The Bacteroides fragilis pathogenicity island is contained in a putative novel conjugative transposon. *J Bacteriol* **186**, 6077-6092.

Fraser, C. M., Read, T.D., Nelson, K.E. (2004). Microbial Genomes. In *Infectious Disease*. Edited by V. St. Georgiev. Totowa, New Jersey: Humana Press.

Fraser, J. D. & Proft, T. (2008). The bacterial superantigen and superantigen-like proteins. *Immunol Rev* **225**, 226-243.

Freestone, P. P., Lyte, M., Neal, C. P., Maggs, A. F., Haigh, R. D. & Williams, P. H. (2000). The mammalian neuroendocrine hormone norepinephrine supplies iron for bacterial growth in the presence of transferrin or lactoferrin. *J Bacteriol* **182**, 6091-6098.

Frick, I. M., Morgelin, M. & Bjorck, L. (2000). Virulent aggregates of Streptococcus pyogenes are generated by homophilic protein-protein interactions. *Mol Microbiol* **37**, 1232-1247.

Frishman, D., Mironov, A., Mewes, H. W. & Gelfand, M. (1998). Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res* **26**, 2941-2947.

Furrer, J. L., Sanders, D. N., Hook-Barnard, I. G. & McIntosh, M. A. (2002). Export of the siderophore enterobactin in Escherichia coli: involvement of a 43 kDa membrane exporter. *Mol Microbiol* **44**, 1225-1234.

Gaillard, J. L., Berche, P., Frehel, C., Gouin, E. & Cossart, P. (1991). Entry of L. monocytogenes into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci. *Cell* **65**, 1127-1141.

Galan, J. E. & Timoney, J. F. (1985a). Mucosal nasopharyngeal immune responses of horses to protein antigens of Streptococcus equi. *Infect Immun* **47**, 623-628.

Galan, J. E. & Timoney, J. F. (1985b). Immune complexes in purpura hemorrhagica of the horse contain IgA and M antigen of Streptococcus equi. *J Immunol* **135**, 3134-3137.

Galan, J. E. & Timoney, J. F. (1987). Molecular analysis of the M protein of Streptococcus equi and cloning and expression of the M protein gene in Escherichia coli. *Infect Immun* **55**, 3181-3187.

Galan, J. E. & Timoney, J. F. (1988). Immunologic and genetic comparison of Streptococcus equi isolates from the United States and Europe. *J Clin Microbiol* **26**, 1142-1146.

Garnier, F., Taourit, S., Glaser, P., Courvalin, P. & Galimand, M. (2000). Characterization of transposon Tn*1549*, conferring VanB-type resistance in *Enterococcus* spp. *Microbiology* **146 (Pt 6)**, 1481-1489.

Gaspar, A. H. & Ton-That, H. (2006). Assembly of distinct pilus structures on the surface of Corynebacterium diphtheriae. *J Bacteriol* **188**, 1526-1533.

Gehring, A. M., Bradley, K. A. & Walsh, C. T. (1997). Enterobactin biosynthesis in Escherichia coli: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36**, 8495-8503.

Gehring, A. M., DeMoll, E., Fetherston, J. D., Mori, I., Mayhew, G. F., Blattner, F. R., Walsh, C. T. & Perry, R. D. (1998a). Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by Yersinia pestis. *Chem Biol* **5**, 573-586.

Gehring, A. M., Mori, I., Perry, R. D. & Walsh, C. T. (1998b). The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of Yersinia pestis. *Biochemistry* **37**, 11637-11650.

Gehring, A. M., Mori, I. & Walsh, C. T. (1998c). Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**, 2648-2659.

George, J. L., Reif, J. S., Shideler, R. K., Small, C. J., Ellis, R. P., Snyder, S. P. & McChesney, A. E. (1983). Identification of carriers of Streptococcus equi in a naturally infected herd. *J Am Vet Med Assoc* **183**, 80-84.

Gianfaldoni, C., Censini, S., Hilleringmann, M. & other authors (2007). Streptococcus pneumoniae pilus subunits protect mice against lethal challenge. *Infect Immun* **75**, 1059-1062.

Giles, R. C., Donahue, J. M., Hong, C. B. & other authors (1993). Causes of abortion, stillbirth, and perinatal death in horses: 3,527 cases (1986-1991). *J Am Vet Med Assoc* **203**, 1170-1175.

Giovanetti, E., Magi, G., Brenciani, A., Spinaci, C., Lupidi, R., Facinelli, B. & Varaldo, P. E. (2002). Conjugative transfer of the erm(A) gene from erythromycin-resistant Streptococcus pyogenes to macrolide-susceptible S. pyogenes, Enterococcus faecalis and Listeria innocua. *J Antimicrob Chemother* **50**, 249-252.

Glaser, P., Rusniok, C., Buchrieser, C. & other authors (2002). Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease. *Mol Microbiol* **45**, 1499-1513.

Grant, S. T., Efstratiou, A. & Chanter, N. (1993). Laboratory diagnosis of strangles and the isolation of atypical Streptococcus equi. *Vet Rec* **133**, 215-216.

Green, N. M., Zhang, S., Porcella, S. F., Nagiec, M. J., Barbian, K. D., Beres, S. B., LeFebvre, R. B. & Musser, J. M. (2005). Genome sequence of a serotype M28 strain of group a streptococcus: potential new insights into puerperal sepsis and bacterial disease specificity. *J Infect Dis* **192**, 760-770.

Grindley, N. D., Whiteson, K. L. & Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**, 567-605.

Gronbaek, L. M., Angen, O., Vigre, H. & Olsen, S. N. (2006). Evaluation of a nested PCR test and bacterial culture of swabs from the nasal passages and from abscesses in relation to diagnosis of Streptococcus equi infection (strangles). *Equine Vet J* **38**, 59-63.

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Res* **36**, W70-74.

Gruenewald, S., Mootz, H. D., Stehmeier, P. & Stachelhaus, T. (2004). In vivo production of artificial nonribosomal peptide products in the heterologous host Escherichia coli. *Appl Environ Microbiol* **70**, 3282-3291.

Guedon, E. & Helmann, J. D. (2003). Origins of metal ion selectivity in the DtxR/MntR family of metalloregulators. *Mol Microbiol* **48**, 495-506.

Hakansson, A., Roche, H., Mirza, S., McDaniel, L. S., Brooks-Walter, A. & Briles, D. E. (2001). Characterization of binding of human lactoferrin to pneumococcal surface protein A. *Infect Immun* **69**, 3372-3381.

Halle, F. & Meyer, J. M. (1992). Iron release from ferrisiderophores. A multi-step mechanism involving a NADH/FMN oxidoreductase and a chemical reduction by FMNH2. *Eur J Biochem* **209**, 621-627.

Hamilton, A., Harrington, D. & Sutcliffe, I. C. (2000). Characterization of acid phosphatase activities in the equine pathogen Streptococcus equi. *Syst Appl Microbiol* **23**, 325-329.

Hamilton, A., Robinson, C., Sutcliffe, I. C., Slater, J., Maskell, D. J., Davis-Poynter, N., Smith, K., Waller, A. & Harrington, D. J. (2006). Mutation of the maturase lipoprotein attenuates the virulence of Streptococcus equi to a greater extent than does loss of general lipoprotein lipidation. *Infect Immun* **74**, 6907-6919.

Hamlen, H. J., Timoney, J. F. & Bell, R. J. (1994). Epidemiologic and immunologic characteristics of Streptococcus equi infection in foals. *J Am Vet Med Assoc* **204**, 768-775.

Han, R., Caswell, C. C., Lukomska, E., Keene, D. R., Pawlowski, M., Bujnicki, J. M., Kim, J. K. & Lukomski, S. (2006a). Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of Streptococcus pyogenes. *Mol Microbiol* **61**, 351-367.

Han, R., Zwiefka, A., Caswell, C. C., Xu, Y., Keene, D. R., Lukomska, E., Zhao, Z., Hook, M. & Lukomski, S. (2006b). Assessment of prokaryotic collagen-like sequences

derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. *Appl Microbiol Biotechnol* **72**, 109-115.

Hanks, T. S., Liu, M., McClure, M. J. & Lei, B. (2005). ABC transporter FtsABCD of Streptococcus pyogenes mediates uptake of ferric ferrichrome. *BMC Microbiol* **5**, 62.

Hantke, K. (2001). Iron and metal regulation in bacteria. *Curr Opin Microbiol* **4**, 172-177.

Hardham, J. M., Stamm, L. V., Porcella, S. F. & other authors (1997). Identification and transcriptional analysis of a Treponema pallidum operon encoding a putative ABC transport system, an iron-activated repressor protein homolog, and a glycolytic pathway enzyme homolog. *Gene* **197**, 47-64.

Harrington, D. J., Greated, J. S., Chanter, N. & Sutcliffe, I. C. (2000). Identification of lipoprotein homologues of pneumococcal PsaA in the equine pathogens Streptococcus equi and Streptococcus zooepidemicus. *Infect Immun* **68**, 6048-6051.

Harrington, D. J., Sutcliffe, I. C. & Chanter, N. (2002). The molecular basis of Streptococcus equi infection and disease. *Microbes Infect* **4**, 501-510.

Hashikawa, S., Iinuma, Y., Furushita, M., Ohkura, T., Nada, T., Torii, K., Hasegawa, T. & Ohta, M. (2004). Characterization of group C and G streptococcal strains that cause streptococcal toxic shock syndrome. *J Clin Microbiol* **42**, 186-192.

Hauck, C. R. & Ohlsen, K. (2006). Sticky connections: extracellular matrix protein recognition and integrin-mediated cellular invasion by Staphylococcus aureus. *Curr Opin Microbiol* **9**, 5-11.

Hava, D. L. & Camilli, A. (2002). Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. *Mol Microbiol* **45**, 1389-1406.

Hayashi, T., Makino, K., Ohnishi, M. & other authors (2001). Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* **8**, 11-22.

Heather, Z., Holden, M. T., Steward, K. F., Parkhill, J., Song, L., Challis, G. L., Robinson, C., Davis-Poynter, N. & Waller, A. S. (2008). A novel streptococcal integrative conjugative element involved in iron acquisition. *Mol Microbiol* **70**, 1274-1292.

Heino, J. (2007). The collagen family members as cell adhesion proteins. *Bioessays* **29**, 1001-1010.

Herwald, H., Cramer, H., Morgelin, M., Russell, W., Sollenberg, U., Norrby-Teglund, A., Flodgaard, H., Lindbom, L. & Bjorck, L. (2004). M protein, a classical bacterial virulence determinant, forms complexes with fibrinogen that induce vascular leakage. *Cell* **116**, 367-379.

Hill, H. R., Bohnsack, J. F., Morris, E. Z., Augustine, N. H., Parker, C. J., Cleary, P. P. & Wu, J. T. (1988). Group B streptococci inhibit the chemotactic activity of the fifth component of complement. *J Immunol* **141**, 3551-3556.

Hoffman, A. M., Staempfli, H. R., Prescott, J. F. & Viel, L. (1991). Field evaluation of a commercial M-protein vaccine against Streptococcus equi infection in foals. *Am J Vet Res* **52**, 589-592.

Hoffman, A. M., Viel, L., Prescott, J. F., Rosendal, S. & Thorsen, J. (1993). Association of microbiologic flora with clinical, endoscopic, and pulmonary cytologic findings in foals with distal respiratory tract infection. *Am J Vet Res* **54**, 1615-1622.

Hofle, G. (2009). The epothilones: an outstanding family of anti-tumor agents. General aspects. *Fortschr Chem Org Naturst* **90**, 5-28.

Holden, M. T., Feil, E. J., Lindsay, J. A. & other authors (2004). Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A* **101**, 9786-9791.

Holden, M. T., Scott, A., Cherevach, I. & other authors (2007). Complete genome of acute rheumatic fever-associated serotype M5 Streptococcus pyogenes strain manfredo. *J Bacteriol* **189**, 1473-1477.

Holden, M. T. G., Heather, Z., Paillot, R., Steward, K. F., Webb, K., Ainslie, F., Jourdan, T., Bason, N. C., Holroyd, N. E., Mungall, K., Quail, M. A., Sanders, M., Simmonds, M., Willey, D., Brooks, K., Aanensen, D. M., Spratt, B. G., Jolley, K. A., Maiden, M. C. J., Kehoe, M., Chanter, N., Bentley, S. D., Robinson, C., Maskell, D. J., Parkhill, J., Waller, A. S. (2009). Genomic Evidence for the Evolution of *Streptococcus equi*: Host Restriction, Increased Virulence and Genetic Exchange with Human Pathogens. *PloS Pathogens* **5**, e1000346.

Holland, R. E. (2008). Practical applications of using PCR testing to eliminate strangles from herds in the USA. In *Getting to Grips with Strangles*. Edinburgh, UK: R & W Communications.

Hollingshead, S. K., Arnold, J., Readdy, T. L. & Bessen, D. E. (1994). Molecular evolution of a multigene family in group A streptococci. *Mol Biol Evol* **11**, 208-219.

Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley, J. T., Williams, S. T. (1994). *Bergey's manual of Determinative Bacteriology*, Ninth edn: Sans Tache, Williams & Wilkins; A Waverley company.

Hong, C. B., Donahue, J. M., Giles, R. C., Jr. & other authors (1993). Etiology and pathology of equine placentitis. *J Vet Diagn Invest* **5**, 56-63.

Hong, K. (2005). Identification and characterization of a novel fibronectin-binding protein gene from Streptococcus equi subspecies zooepidemicus strain VTU211. *FEMS Immunol Med Microbiol* **45**, 231-237.

Hoskins, J., Alborn, W. E., Jr., Arnold, J. & other authors (2001). Genome of the bacterium Streptococcus pneumoniae strain R6. *J Bacteriol* **183**, 5709-5717.

Huang, I.-H., Wu, C., Merritt, J., Qi, F. (2008). Transcriptome Pofiling of Cellular Responses of *Streptococcus mutans* and *Streptococcus gordonii* to Interspecies Antagonism in the Oral biofilm. In *XVII Lancefield International Symposium on Streptococci and Streptococcal diseases*. Porto Heli, Greece.

Humtsoe, J. O., Kim, J. K., Xu, Y., Keene, D. R., Hook, M., Lukomski, S. & Wary, K. K. (2005). A streptococcal collagen-like protein interacts with the alpha2beta1 integrin and induces intracellular signaling. *J Biol Chem* **280**, 13848-13857.

Hynes, W. L. & Walton, S. L. (2000). Hyaluronidases of Gram-positive bacteria. *FEMS Microbiol Lett* **183**, 201-207.

Ikebe, T., Wada, A., Inagaki, Y. & other authors (2002). Dissemination of the phage-associated novel superantigen gene *speL* in recent invasive and noninvasive *Streptococcus pyogenes* M3/T3 isolates in Japan. *Infect Immun* **70**, 3227-3233.

Ireton, K. (2007). Entry of the bacterial pathogen *Listeria monocytogenes* into mammalian cells. *Cell Microbiol* **9**, 1365-1375.

Jacobs, A. A., Goovaerts, D., Nuijten, P. J., Theelen, R. P., Hartford, O. M. & Foster, T. J. (2000). Investigations towards an efficacious and safe strangles vaccine: submucosal vaccination with a live attenuated Streptococcus equi. *Vet Rec* **147**, 563-567.

Janulczyk, R., Pallon, J. & Bjorck, L. (1999). Identification and characterization of a Streptococcus pyogenes ABC transporter with multiple specificity for metal cations. *Mol Microbiol* **34**, 596-606.

Janulczyk, R., Ricci, S. & Bjorck, L. (2003). MtsABC is important for manganese and iron transport, oxidative stress resistance, and virulence of Streptococcus pyogenes. *Infect Immun* **71**, 2656-2664.

Jeffcott, L. B., Rossdale, P. D., Freestone, J., Frank, C. J. & Towers-Clark, P. F. (1982). An assessment of wastage in thoroughbred racing from conception to 4 years of age. *Equine Vet J* **14**, 185-198.

Ji, Y., McLandsborough, L., Kondagunta, A. & Cleary, P. P. (1996). C5a peptidase alters clearance and trafficking of group A streptococci by infected mice. *Infect Immun* **64**, 503-510.

Jin, Q., Yuan, Z., Xu, J. & other authors (2002). Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. *Nucleic Acids Res* **30**, 4432-4441.

Jolley, K. A., Chan, M. S. & Maiden, M. C. (2004). mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* **5**, 86.

Jonsson, H., Lindmark, H. & Guss, B. (1995). A protein G-related cell surface protein in Streptococcus zooepidemicus. *Infect Immun* **63**, 2968-2975.

Jorm, L. R. (1990). Strangles in horse studs: incidence, risk factors and effect of vaccination. *Aust Vet J* **67**, 436-439.

Jorm, L. R. (1992). Laboratory studies on the survival of *Streptococcus equi* subspecies *equi* on surfaces. In *Proceedings of Equine Infectious Diseases VI*, pp. 39-43. Newmarket, UK: R & W Publications Ltd.

Jorm, L. R., Love, D. N., Bailey, G. D., McKay, G. M. & Briscoe, D. A. (1994). Genetic structure of populations of beta-haemolytic Lancefield group C streptococci from horses and their association with disease. *Res Vet Sci* **57**, 292-299.

Judy, C. E., Chaffin, M. K. & Cohen, N. D. (1999). Empyema of the guttural pouch (auditory tube diverticulum) in horses: 91 cases (1977-1997). *J Am Vet Med Assoc* **215**, 1666-1670.

Kang, H. J., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. (2007). Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science* **318**, 1625-1628.

Karlstrom, A., Jacobsson, K., Flock, M., Flock, J. I. & Guss, B. (2004). Identification of a novel collagen-like protein, SclC, in Streptococcus equi using signal sequence phage display. *Vet Microbiol* **104**, 179-188.

Karlstrom, A., Jacobsson, K. & Guss, B. (2006). SclC is a member of a novel family of collagen-like proteins in Streptococcus equi subspecies equi that are recognised by antibodies against SclC. *Vet Microbiol* **114**, 72-81.

Keating, T. A., Miller, D. A. & Walsh, C. T. (2000). Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of Yersiniabactin synthetase. *Biochemistry* **39**, 4729-4739.

Kelly, C., Bugg, M., Robinson, C., Mitchell, Z., Davis-Poynter, N., Newton, J. R., Jolley, K. A., Maiden, M. C. & Waller, A. S. (2006). Sequence variation of the SeM gene of Streptococcus equi allows discrimination of the source of strangles outbreaks. *J Clin Microbiol* **44**, 480-486.

Kemp-Symonds, J., Kemble, T. & Waller, A. (2007). Modified live Streptococcus equi ('strangles') vaccination followed by clinically adverse reactions associated with bacterial replication. *Equine Vet J* **39**, 284-286.

Kleinkauf, H. & von Dohren, H. (1990). Nonribosomal biosynthesis of peptide antibiotics. *Eur J Biochem* **192**, 1-15.

Klumpp, C., Burger, A., Mislin, G. L. & Abdallah, M. A. (2005). From a total synthesis of cepabactin and its 3:1 ferric complex to the isolation of a 1:1:1 mixed complex between iron (III), cepabactin and pyochelin. *Bioorg Med Chem Lett* **15**, 1721-1724.

Konz, D., Klens, A., Schorgendorfer, K. & Marahiel, M. A. (1997). The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* **4**, 927-937.

Konz, D. & Marahiel, M. A. (1999). How do peptide synthetases generate structural diversity? *Chem Biol* **6**, R39-48.

Korman, T. M., Boers, A., Gooding, T. M., Curtis, N. & Visvanathan, K. (2004). Fatal case of toxic shock-like syndrome due to group C streptococcus associated with superantigen exotoxin. *J Clin Microbiol* **42**, 2866-2869.

Koster, W. & Braun, V. (1990). Iron (III) hydroxamate transport into Escherichia coli. Substrate binding to the periplasmic FhuD protein. *J Biol Chem* **265**, 21407-21410.

Kotb, M. (1995). Bacterial pyrogenic exotoxins as superantigens. *Clin Microbiol Rev* **8**, 411-426.

Krishnan, V., Gaspar, A. H., Ye, N., Mandlik, A., Ton-That, H. & Narayana, S. V. (2007). An IgG-like domain in the minor pilin GBS52 of Streptococcus agalactiae mediates lung epithelial cell adhesion. *Structure* **15**, 893-903.

Kunin, V., Sorek, R. & Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* **8**, R61.

Kuusi, M., Lahti, E., Virolainen, A. & other authors (2006). An outbreak of Streptococcus equi subspecies zooepidemicus associated with consumption of fresh goat cheese. *BMC Infect Dis* **6**, 36.

Ladlow, J., Scase, T. & Waller, A. (2006). Canine strangles case reveals a new host susceptible to infection with Streptococcus equi. *J Clin Microbiol* **44**, 2664-2665.

Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C. & Walsh, C. T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **3**, 923-936.

Lancefield, R. (1933). A serological differentiation of human and other groups of streptococci. *J Exp Med* **57**, 571-595.

Langermann, S. & Ballou, W. R., Jr. (2001). Vaccination utilizing the FimCH complex as a strategy to prevent Escherichia coli urinary tract infections. *J Infect Dis* **183 Suppl 1**, S84-86.

Lanie, J. A., Ng, W. L., Kazmierczak, K. M., Andrzejewski, T. M., Davidsen, T. M., Wayne, K. J., Tettelin, H., Glass, J. I. & Winkler, M. E. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6. *J Bacteriol* **189**, 38-51.

Lannergard, J., Frykberg, L. & Guss, B. (2003). CNE, a collagen-binding protein of *Streptococcus equi*. *FEMS Microbiol Lett* **222**, 69-74.

Lannergard, J., Flock, M., Johansson, S., Flock, J. I. & Guss, B. (2005). Studies of fibronectin-binding proteins of Streptococcus equi. *Infect Immun* **73**, 7243-7251.

Lannergard, J. (2006). Potentially Virulence-Related Extracellular Proteins of *Streptococcus equi*. In *Faculty of Natural Resources and Agricultural Sciences*. Uppsala: Swedish University of Agricultural Sciences.

Lannergard, J. & Guss, B. (2006). IdeE, an IgG-endopeptidase of Streptococcus equi ssp. equi. *FEMS Microbiol Lett* **262**, 230-235.

Las Heras, A., Vela, A. I., Fernandez, E., Legaz, E., Dominguez, L. & Fernandez-Garayzabal, J. F. (2002). Unusual outbreak of clinical mastitis in dairy sheep caused by Streptococcus equi subsp. zooepidemicus. *J Clin Microbiol* **40**, 1106-1108.

Latorre, M., Alvarez, M., Fernandez, J. M., Berdonces, P., Llanos, A. & Cisterna, R. (1993). A case of meningitis due to "Streptococcus zooepidemicus". *Clin Infect Dis* **17**, 932-933.

Lauer, P., Rinaudo, C. D., Soriani, M. & other authors (2005). Genome analysis reveals pili in Group B Streptococcus. *Science* **309**, 105.

Lautru, S. & Challis, G. L. (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology* **150**, 1629-1636.

Lautru, S., Deeth, R. J., Bailey, L. M. & Challis, G. L. (2005). Discovery of a new peptide natural product by Streptomyces coelicolor genome mining. *Nat Chem Biol* **1**, 265-269.

Lei, B., DeLeo, F. R., Reid, S. D. & other authors (2002a). Opsonophagocytosis-inhibiting mac protein of group a streptococcus: identification and characteristics of two genetic complexes. *Infect Immun* **70**, 6880-6890.

Lei, B., Smoot, L. M., Menning, H. M., Voyich, J. M., Kala, S. V., Deleo, F. R., Reid, S. D. & Musser, J. M. (2002b). Identification and characterization of a novel heme-associated cell surface protein made by Streptococcus pyogenes. *Infect Immun* **70**, 4494-4500.

Lei, B., Liu, M., Meyers, E. G., Manning, H. M., Nagiec, M. J. & Musser, J. M. (2003). Histidine and aspartic acid residues important for immunoglobulin G endopeptidase activity of the group A Streptococcus opsonophagocytosis-inhibiting Mac protein. *Infect Immun* **71**, 2881-2884.

LeMieux, J., Woody, S. & Camilli, A. (2008). Roles of the sortases of Streptococcus pneumoniae in assembly of the RlrA pilus. *J Bacteriol* **190**, 6002-6013.

Lewis, M. J., Meehan, M., Owen, P. & Woof, J. M. (2008a). A common theme in interaction of bacterial immunoglobulin-binding proteins with immunoglobulins illustrated in the equine system. *J Biol Chem* **283**, 17615-17623.

Lewis, M. J., Wagner, B. & Woof, J. M. (2008b). The different effector function capabilities of the seven equine IgG subclasses have implications for vaccine strategies. *Mol Immunol* **45**, 818-827.

Li, J. J. & Bugg, T. D. (2007). Investigation of a general base mechanism for ester hydrolysis in C-C hydrolase enzymes of the alpha/beta-hydrolase superfamily: a novel mechanism for the serine catalytic triad. *Org Biomol Chem* **5**, 507-513.

Li, X. Z., Nikaido, H. & Poole, K. (1995). Role of mexA-mexB-oprM in antibiotic efflux in Pseudomonas aeruginosa. *Antimicrob Agents Chemother* **39**, 1948-1953.

Liden, A., Karlstrom, A., Lannergard, J., Kalamajski, S., Guss, B., Rubin, K. & Ryden, C. (2006). A fibronectin-binding protein from Streptococcus equi binds collagen and modulates cell-mediated collagen gel contraction. *Biochem Biophys Res Commun* **340**, 604-610.

Liden, A., van Wieringen, T., Lannergard, J., Kassner, A., Heinegard, D., Reed, R. K., Guss, B. & Rubin, K. (2008). A secreted collagen- and fibronectin-binding streptococcal protein modulates cell-mediated collagen gel contraction and interstitial fluid pressure. *J Biol Chem* **283**, 1234-1242.

Lindmark, H., Jacobsson, K., Frykberg, L. & Guss, B. (1996). Fibronectin-binding protein of *Streptococcus equi* subsp. *zooepidemicus*. *Infect Immun* **64**, 3993-3999.

Lindmark, H. & Guss, B. (1999). SFS, a novel fibronectin-binding protein from *Streptococcus equi*, inhibits the binding between fibronectin and collagen. *Infect Immun* **67**, 2383-2388.

Lindmark, H., Jonsson, P., Engvall, E. & Guss, B. (1999). Pulsed-field gel electrophoresis and distribution of the genes *zag* and *fnz* in isolates of *Streptococcus equi*. *Res Vet Sci* **66**, 93-99.

Lindmark, H., Nilsson, M. & Guss, B. (2001). Comparison of the fibronectin-binding protein FNE from *Streptococcus equi* subspecies *equi* with FNZ from *S. equi* subspecies *zooepidemicus* reveals a major and conserved difference. *Infect Immun* **69**, 3159-3163.

Lindsay, A. M., Zhang, M., Mitchell, Z., Holden, M. T., Waller, A. S., Sutcliffe, I. C. & Black, G. W. (2009). The Streptococcus equi prophage-encoded protein SEQ2045 is a hyaluronan-specific hyaluronate lyase that is produced during equine infection. *Microbiology* **155**, 443-449.

Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. (1998). The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in Staphylococcus aureus. *Mol Microbiol* **29**, 527-543.

Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P. & Moras, D. (1998). Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell* **95**, 771-778.

Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Ireland, R. M., Reid, S. D., Adams, G. G. & Musser, J. M. (2000). Identification and characterization of the scl gene encoding a group A Streptococcus extracellular protein virulence factor with similarity to human collagen. *Infect Immun* **68**, 6542-6553.

Maione, D., Margarit, I., Rinaudo, C. D. & other authors (2005). Identification of a universal Group B *Streptococcus* vaccine by multiple genome screen. *Science* **309**, 148-150.

Maisey, H. C., Hensler, M., Nizet, V. & Doran, K. S. (2007). Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells. *J Bacteriol* **189**, 1464-1467.

Mandlik, A., Swierczynski, A., Das, A. & Ton-That, H. (2007). Corynebacterium diphtheriae employs specific minor pilins to target human pharyngeal epithelial cells. *Mol Microbiol* **64**, 111-124.

Mandlik, A., Swierczynski, A., Das, A. & Ton-That, H. (2008). Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol* **16**, 33-40.

Manetti, A. G., Zingaretti, C., Falugi, F. & other authors (2007). *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. *Mol Microbiol* **64**, 968-983.

Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev* **97**, 2651-2674.

Marciel, A. M., Kapur, V. & Musser, J. M. (1997). Molecular population genetic analysis of a Streptococcus pyogenes bacteriophage-encoded hyaluronidase gene: recombination contributes to allelic variation. *Microb Pathog* **22**, 209-217.

Marimon, J. M., Valiente, A., Ercibengoa, M., Garcia-Arenzana, J. M. & Perez-Trallero, E. (2005). Erythromycin resistance and genetic elements carrying macrolide efflux genes in Streptococcus agalactiae. *Antimicrob Agents Chemother* **49**, 5069-5074.

Marraffini, L. A., Dedent, A. C. & Schneewind, O. (2006). Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* **70**, 192-221.

Maurelli, A. T. (2007). Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiol Lett* **267**, 1-8.

May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. (2002). Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* **99**, 12120-12125.

May, J. P., Walker, C. A., Maskell, D. J. & Slater, J. D. (2004). Development of an in vivo Himar1 transposon mutagenesis system for use in Streptococcus equi subsp. equi. *FEMS Microbiol Lett* **238**, 401-409.

McCoy, H. E., Broder, C. C. & Lottenberg, R. (1991). Streptokinases produced by pathogenic group C streptococci demonstrate species-specific plasminogen activation. *J Infect Dis* **164**, 515-521.

McElroy, M. C., Cain, D. J., Tyrrell, C., Foster, T. J. & Haslett, C. (2002). Increased virulence of a fibronectin-binding protein mutant of *Staphylococcus aureus* in a rat model of pneumonia. *Infect Immun* **70**, 3865-3873.

McGorum, B. C., Robinson, N. E., Schumacher, J., Dixon, P. M. (2006). Equine Respiratory Medicine and Surgery: Saunders Ltd.

McShan, W. M., Ferretti, J. J., Karasawa, T. & other authors (2008). Genome Sequence of a Nephritogenic and Highly Transformable M49 Strain of Streptococcus pyogenes. *J Bacteriol.*

Meehan, M., Nowlan, P. & Owen, P. (1998). Affinity purification and characterization of a fibrinogen-binding protein complex which protects mice against lethal challenge with Streptococcus equi subsp. equi. *Microbiology* **144 (Pt 4)**, 993-1003.

Meehan, M., Lynagh, Y., Woods, C. & Owen, P. (2001). The fibrinogen-binding protein (FgBP) of Streptococcus equi subsp. equi additionally binds IgG and contributes to virulence in a mouse model. *Microbiology* **147**, 3311-3322.

Meehan, M., Kelly, S. M., Price, N. C. & Owen, P. (2002). The C-terminal portion of the fibrinogen-binding protein of Streptococcus equi subsp. equi contains extensive alpha-helical coiled-coil structure and contributes to thermal stability. *FEMS Microbiol Lett* **206**, 81-86.

Meeusen, E. N., Walker, J., Peters, A., Pastoret, P. P. & Jungersen, G. (2007). Current status of veterinary vaccines. *Clin Microbiol Rev* **20**, 489-510, table of contents.

Michel, L., Gonzalez, N., Jagdeep, S., Nguyen-Ngoc, T. & Reimmann, C. (2005). PchR-box recognition by the AraC-type regulator PchR of Pseudomonas aeruginosa requires the siderophore pyochelin as an effector. *Mol Microbiol* **58**, 495-509.

Michel, L., Bachelard, A. & Reimmann, C. (2007). Ferripyochelin uptake genes are involved in pyochelin-mediated signalling in Pseudomonas aeruginosa. *Microbiology* **153**, 1508-1518.

Miethke, M. & Marahiel, M. A. (2007). Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev* **71**, 413-451.

Miller, D. A., Luo, L., Hillson, N., Keating, T. A. & Walsh, C. T. (2002). Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem Biol* **9**, 333-344.

Mora, M., Bensi, G., Capo, S. & other authors (2005). Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc Natl Acad Sci U S A* **102**, 15641-15646.

Moschioni, M., Donati, C., Muzzi, A. & other authors (2008). Streptococcus pneumoniae contains 3 rlrA pilus variants that are clonally related. *J Infect Dis* **197**, 888-896.

Mousel, M. R., Harrison, L., Donahue, J. M. & Bailey, E. (2003). Rhodococcus equi and genetic susceptibility: assessing transferrin genotypes from paraffin-embedded tissues. *J Vet Diagn Invest* **15**, 470-472.

Nagiec, M. J., Lei, B., Parker, S. K., Vasil, M. L., Matsumoto, M., Ireland, R. M., Beres, S. B., Hoe, N. P. & Musser, J. M. (2004). Analysis of a novel prophage-encoded group A Streptococcus extracellular phospholipase A(2). *J Biol Chem* **279**, 45909-45918.

Nakagawa, I., Kurokawa, K., Yamashita, A. & other authors (2003). Genome sequence of an M3 strain of Streptococcus pyogenes reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Res* **13**, 1042-1055.

Nakata, N., Tobe, T., Fukuda, I., Suzuki, T., Komatsu, K., Yoshikawa, M. & Sasakawa, C. (1993). The absence of a surface protease, OmpT, determines the intercellular spreading ability of Shigella: the relationship between the ompT and kcpA loci. *Mol Microbiol* **9**, 459-468.

Nallapareddy, S. R., Weinstock, G. M. & Murray, B. E. (2003). Clinical isolates of Enterococcus faecium exhibit strain-specific collagen binding mediated by Acm, a new member of the MSCRAMM family. *Mol Microbiol* **47**, 1733-1747.

Nassif, X., Lowy, J., Stenberg, P., O'Gaora, P., Ganji, A. & So, M. (1993). Antigenic variation of pilin regulates adhesion of Neisseria meningitidis to human epithelial cells. *Mol Microbiol* **8**, 719-725.

Nelson, A. L., Ries, J., Bagnoli, F. & other authors (2007). RrgA is a pilus-associated adhesin in Streptococcus pneumoniae. *Mol Microbiol* **66**, 329-340.

Newton, J. R., Wood, J. L., Dunn, K. A., DeBrauwere, M. N. & Chanter, N. (1997). Naturally occurring persistent and asymptomatic infection of the guttural pouches of horses with Streptococcus equi. *Vet Rec* **140**, 84-90.

Newton, J. R., Verheyen, K., Talbot, N. C., Timoney, J. F., Wood, J. L., Lakhani, K. H. & Chanter, N. (2000). Control of strangles outbreaks by isolation of guttural pouch carriers identified using PCR and culture of Streptococcus equi. *Equine Vet J* **32**, 515-526.

Newton, J. R. (2002). Epidemiological studies of inflammatory airway disease in horses. In *Life Sciences*: Open University.

Newton, J. R. (2003). Investigations of respiratory diseases of the Thoroughbred racehorse: RCVS Fellowship Thesis.

Newton, J. R., Wood, J. L. & Chanter, N. (2003). A case control study of factors and infections associated with clinically apparent respiratory disease in UK Thoroughbred racehorses. *Prev Vet Med* **60**, 107-132.

Newton, J. R., Woodt, J. L. & Chanter, N. (2007). Evidence for transferrin allele as a host-level risk factor in naturally occurring equine respiratory disease: a preliminary study. *Equine Vet J* **39**, 164-171.

Newton, J. R., Laxton, R., Wood, J. L. & Chanter, N. (2008). Molecular epidemiology of Streptococcus zooepidemicus infection in naturally occurring equine respiratory disease. *Vet J* **175**, 338-345.

Nickel, V., Prehm, S., Lansing, M., Mausolf, A., Podbielski, A., Deutscher, J. & Prehm, P. (1998). An ectoprotein kinase of group C streptococci binds hyaluronan and regulates capsule formation. *J Biol Chem* **273**, 23668-23673.

Nierman, W. C., DeShazer, D., Kim, H. S. & other authors (2004). Structural flexibility in the *Burkholderia mallei* genome. *Proc Natl Acad Sci U S A* **101**, 14246-14251.

Nyberg, P., Sakai, T., Cho, K. H., Caparon, M. G., Fassler, R. & Bjorck, L. (2004). Interactions with fibronectin attenuate the virulence of Streptococcus pyogenes. *EMBO J* **23**, 2166-2174.

Nygaard, T. K., Liu, M., McClure, M. J. & Lei, B. (2006). Identification and characterization of the heme-binding proteins SeShp and SeHtsA of Streptococcus equi subspecies equi. *BMC Microbiol* **6**, 82.

Oikawa, M., Kamada, M., Yoshikawa, Y. & Yoshikawa, T. (1994). Pathology of equine pneumonia associated with transport and isolation of Streptococcus equi subsp. zooepidemicus. *J Comp Pathol* **111**, 205-212.

Oram, D. M., Must, L. M., Spinler, J. K., Twiddy, E. M. & Holmes, R. K. (2005). Analysis of truncated variants of the iron dependent transcriptional regulators from Corynebacterium diphtheriae and Mycobacterium tuberculosis. *FEMS Microbiol Lett* **243**, 1-8.

Orsi, R. H., Ripoll, D. R., Yeung, M., Nightingale, K. K. & Wiedmann, M. (2007). Recombination and positive selection contribute to evolution of *Listeria monocytogenes* inlA. *Microbiology* **153**, 2666-2678.

Osorio, C. R., Juiz-Rio, S. & Lemos, M. L. (2006). A siderophore biosynthesis gene cluster from the fish pathogen Photobacterium damselae subsp. piscicida is structurally and functionally related to the Yersinia high-pathogenicity island. *Microbiology* **152**, 3327-3341.

Otten, L. G., Schaffer, M. L., Villiers, B. R., Stachelhaus, T. & Hollfelder, F. (2007). An optimized ATP/PP(i)-exchange assay in 96-well format for screening of adenylation domains for applications in combinatorial biosynthesis. *Biotechnol J* **2**, 232-240.

Page, W. J., Kwon, E., Cornish, A. S. & Tindale, A. E. (2003). The csbX gene of Azotobacter vinelandii encodes an MFS efflux pump required for catecholate siderophore export. *FEMS Microbiol Lett* **228**, 211-216.

Pahlman, L. I., Morgelin, M., Eckert, J. & other authors (2006). Streptococcal M protein: a multipotent and powerful inducer of inflammation. *J Immunol* **177**, 1221-1228.

Pahlman, L. I., Marx, P. F., Morgelin, M., Lukomski, S., Meijers, J. C. & Herwald, H. (2007). Thrombin-activatable fibrinolysis inhibitor binds to Streptococcus pyogenes by interacting with collagen-like proteins A and B. *J Biol Chem* **282**, 24873-24881.

Pallen, M. J. & Wren, B. W. (2007). Bacterial pathogenomics. *Nature* **449**, 835-842.

Parillo, F., Rossi, G., Busoni, V., Magi, G. E. & Verini Supplizi, A. (2009). Differentiation of glycans in equine guttural pouches. *Vet J* **180**, 246-252.

Parkhill, J., Achtman, M., James, K. D. & other authors (2000). Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491. *Nature* **404**, 502-506.

Parkhill, J., Dougan, G., James, K. D. & other authors (2001a). Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-852.

Parkhill, J., Wren, B. W., Thomson, N. R. & other authors (2001b). Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* **413**, 523-527.

Parkhill, J., Sebaihia, M., Preston, A. & other authors (2003). Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nat Genet* **35**, 32-40.

Patel, H. M. & Walsh, C. T. (2001). In vitro reconstitution of the Pseudomonas aeruginosa nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. *Biochemistry* **40**, 9023-9031.

Paterson, G. K., Nieminen, L., Jefferies, J. M. & Mitchell, T. J. (2008). PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol Lett* **285**, 170-176.

Pati, S., Al-Araji, A. & Orendi, J. (2007). Atypical presentation of Streptococcus zooepidemicus bacteraemia and secondary meningitis. *Clin Neurol Neurosurg* **109**, 475-476.

Patti, J. M., Jonsson, H., Guss, B., Switalski, L. M., Wiberg, K., Lindberg, M. & Hook, M. (1992). Molecular characterization and expression of a gene encoding a Staphylococcus aureus collagen adhesin. *J Biol Chem* **267**, 4766-4772.

Patti, J. M., Allen, B. L., McGavin, M. J. & Hook, M. (1994). MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* **48**, 585-617.

Pearce, S. R., Mimmack, M. L., Gallagher, M. P., Gileadi, U., Hyde, S. C. & Higgins, C. F. (1992). Membrane topology of the integral membrane components, OppB and OppC, of the oligopeptide permease of Salmonella typhimurium. *Mol Microbiol* **6**, 47-57.

Pelludat, C., Rakin, A., Jacobi, C. A., Schubert, S. & Heesemann, J. (1998). The yersiniabactin biosynthetic gene cluster of Yersinia enterocolitica: organization and siderophore-dependent regulation. *J Bacteriol* **180**, 538-546.

Perna, N. T., Mayhew, G. F., Posfai, G., Elliott, S., Donnenberg, M. S., Kaper, J. B. & Blattner, F. R. (1998). Molecular evolution of a pathogenicity island from enterohemorrhagic Escherichia coli O157:H7. *Infect Immun* **66**, 3810-3817.

Perna, N. T., Plunkett, G., 3rd, Burland, V. & other authors (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature* **409**, 529-533.

Perry, R. D., Balbo, P. B., Jones, H. A., Fetherston, J. D. & DeMoll, E. (1999). Yersiniabactin from Yersinia pestis: biochemical characterization of the siderophore and its role in iron transport and regulation. *Microbiology* **145 (Pt 5)**, 1181-1190.

Pesavento, P. A., Hurley, K. F., Bannasch, M. J., Artiushin, S. & Timoney, J. F. (2008). A clonal outbreak of acute fatal hemorrhagic pneumonia in intensively housed (shelter) dogs caused by Streptococcus equi subsp. zooepidemicus. *Vet Pathol* **45**, 51-53.

Pezzicoli, A., Santi, I., Lauer, P., Rosini, R., Rinaudo, D., Grandi, G., Telford, J. L. & Soriani, M. (2008). Pilus backbone contributes to group B Streptococcus paracellular translocation through epithelial cells. *J Infect Dis* **198**, 890-898.

Pfeifer, B. A., Wang, C. C., Walsh, C. T. & Khosla, C. (2003). Biosynthesis of Yersiniabactin, a complex polyketide-nonribosomal peptide, using Escherichia coli as a heterologous host. *Appl Environ Microbiol* **69**, 6698-6702.

Pierre, J. L., Fontecave, M. & Crichton, R. R. (2002). Chemistry for an essential biological process: the reduction of ferric iron. *Biometals* **15**, 341-346.

Pohl, E., Holmes, R. K. & Hol, W. G. (1999a). Crystal structure of the iron-dependent regulator (IdeR) from Mycobacterium tuberculosis shows both metal binding sites fully occupied. *J Mol Biol* **285**, 1145-1156.

Pohl, E., Holmes, R. K. & Hol, W. G. (1999b). Crystal structure of a cobalt-activated diphtheria toxin repressor-DNA complex reveals a metal-binding SH3-like domain. *J Mol Biol* **292**, 653-667.

Ponnuraj, K., Bowden, M. G., Davis, S., Gurusiddappa, S., Moore, D., Choe, D., Xu, Y., Hook, M. & Narayana, S. V. (2003). A "dock, lock, and latch" structural model for a staphylococcal adhesin binding to fibrinogen. *Cell* **115**, 217-228.

Pramanik, A. & Braun, V. (2006). Albomycin uptake via a ferric hydroxamate transport system of Streptococcus pneumoniae R6. *J Bacteriol* **188**, 3878-3886.

Pritchard, D. G., Lin, B., Willingham, T. R. & Baker, J. R. (1994). Characterization of the group B streptococcal hyaluronate lyase. *Arch Biochem Biophys* **315**, 431-437.

Proft, T., Moffatt, S. L., Berkahn, C. J. & Fraser, J. D. (1999). Identification and characterization of novel superantigens from Streptococcus pyogenes. *J Exp Med* **189**, 89-102.

Proft, T., Webb, P. D., Handley, V. & Fraser, J. D. (2003). Two novel superantigens found in both group A and group C *Streptococcus*. *Infect Immun* **71**, 1361-1369.

Pusterla, N., Watson, J. L., Affolter, V. K., Magdesian, K. G., Wilson, W. D. & Carlson, G. P. (2003). Purpura haemorrhagica in 53 horses. *Vet Rec* **153**, 118-121.

Qiu, X., Pohl, E., Holmes, R. K. & Hol, W. G. (1996). High-resolution structure of the diphtheria toxin repressor complexed with cobalt and manganese reveals an SH3-like third domain and suggests a possible role of phosphate as co-corepressor. *Biochemistry* **35**, 12292-12302.

Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T. (1998). Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595.

Quadri, L. E., Keating, T. A., Patel, H. M. & Walsh, C. T. (1999). Assembly of the Pseudomonas aeruginosa nonribosomal peptide siderophore pyochelin: In vitro reconstitution of aryl-4, 2-bisthiazoline synthetase activity from PchD, PchE, and PchF. *Biochemistry* **38**, 14941-14954.

Que, Q. & Helmann, J. D. (2000). Manganese homeostasis in Bacillus subtilis is regulated by MntR, a bifunctional regulator related to the diphtheria toxin repressor family of proteins. *Mol Microbiol* **35**, 1454-1468.

Rakonjac, J. V., Robbins, J. C. & Fischetti, V. A. (1995). DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. *Infect Immun* **63**, 622-631.

Ramachandran, V., McArthur, J. D., Behm, C. E. & other authors (2004). Two distinct genotypes of prtF2, encoding a fibronectin binding protein, and evolution of the gene family in Streptococcus pyogenes. *J Bacteriol* **186**, 7601-7609.

Raphel, C. F. & Beech, J. (1982). Pleuritis secondary to pneumonia or lung abscessation in 90 horses. *J Am Vet Med Assoc* **181**, 808-810.

Rasmussen, M., Muller, H. P. & Bjorck, L. (1999). Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin. *J Biol Chem* **274**, 15336-15344.

Rasmussen, M. & Bjorck, L. (2001). Unique regulation of SclB - a novel collagen-like surface protein of Streptococcus pyogenes. *Mol Microbiol* **40**, 1427-1438.

Rasmussen, M. & Bjorck, L. (2002). Proteolysis and its regulation at the surface of Streptococcus pyogenes. *Mol Microbiol* **43**, 537-544.

Rasmussen, M., Jacobsson, M. & Bjorck, L. (2003). Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem* **278**, 32313-32316.

Read, T. D., Satola, S. W. & Farley, M. M. (2000). Nucleotide sequence analysis of hypervariable junctions of Haemophilus influenzae pilus gene clusters. *Infect Immun* **68**, 6896-6902.

Reichardt, W., Gubbe, K. & Schmidt, K. H. (1995). M3-protein with close sequence homology to M12 protein binds fibrinogen, albumin, fibronectin, but not to any subclass of IgG-localization of binding regions. *Dev Biol Stand* **85**, 179-182.

Reig, M., Galan, J., Baquero, F. & Perez-Diaz, J. C. (2001). Macrolide resistance in Peptostreptococcus spp. mediated by ermTR: possible source of macrolide-lincosamide-streptogramin B resistance in Streptococcus pyogenes. *Antimicrob Agents Chemother* **45**, 630-632.

Reimmann, C., Patel, H. M., Serino, L., Barone, M., Walsh, C. T. & Haas, D. (2001). Essential PchG-dependent reduction in pyochelin biosynthesis of Pseudomonas aeruginosa. *J Bacteriol* **183**, 813-820.

Reimmann, C., Patel, H. M., Walsh, C. T. & Haas, D. (2004). PchC thioesterase optimizes nonribosomal biosynthesis of the peptide siderophore pyochelin in Pseudomonas aeruginosa. *J Bacteriol* **186**, 6367-6373.

Rich, R. L., Kreikemeyer, B., Owens, R. T., LaBrenz, S., Narayana, S. V., Weinstock, G. M., Murray, B. E. & Hook, M. (1999). Ace is a collagen-binding MSCRAMM from Enterococcus faecalis. *J Biol Chem* **274**, 26939-26945.

Roberts, A. P., Johanesen, P. A., Lyras, D., Mullany, P. & Rood, J. I. (2001). Comparison of Tn5397 from Clostridium difficile, Tn916 from Enterococcus faecalis and

the CW459tet(M) element from Clostridium perfringens shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology* **147**, 1243-1251.

Robinson, D. A., Sutcliffe, J. A., Tewodros, W., Manoharan, A. & Bessen, D. E. (2006). Evolution and global dissemination of macrolide-resistant group A streptococci. *Antimicrob Agents Chemother* **50**, 2903-2911.

Rodionov, D. A., Hebbeln, P., Gelfand, M. S. & Eitinger, T. (2006). Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J Bacteriol* **188**, 317-327.

Rodionov, D. A., Hebbeln, P., Eudes, A. & other authors (2009). A novel class of modular transporters for vitamins in prokaryotes. *J Bacteriol* **191**, 42-51.

Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. & Smith, I. (2002). ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun* **70**, 3371-3381.

Rodriguez, G. M. & Smith, I. (2006). Identification of an ABC transporter required for iron acquisition and virulence in Mycobacterium tuberculosis. *J Bacteriol* **188**, 424-430.

Rosini, R., Rinaudo, C. D., Soriani, M. & other authors (2006). Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae. *Mol Microbiol* **61**, 126-141.

Rossdale, P. D., Hopes, R., Digby, N. J. & offord, K. (1985). Epidemiological study of wastage among racehorses 1982 and 1983. *Vet Rec* **116**, 66-69.

Ruzin, A., Lindsay, J. & Novick, R. P. (2001). Molecular genetics of SaPI1--a mobile pathogenicity island in Staphylococcus aureus. *Mol Microbiol* **41**, 365-377.

Salama, N., Guillemin, K., McDaniel, T. K., Sherlock, G., Tompkins, L. & Falkow, S. (2000). A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. *Proc Natl Acad Sci U S A* **97**, 14668-14673.

Salasia, S. I., Wibawan, I. W., Pasaribu, F. H., Abdulmawjood, A. & Lammler, C. (2004). Persistent occurrence of a single Streptococcus equi subsp. zooepidemicus clone in the pig and monkey population in Indonesia. *J Vet Sci* **5**, 263-265.

Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26**, 544-548.

Sansonetti, P. J., Kopecko, D. J. & Formal, S. B. (1982). Involvement of a plasmid in the invasive ability of Shigella flexneri. *Infect Immun* **35**, 852-860.

Sato, Y., Okamoto, K., Kagami, A., Yamamoto, Y., Igarashi, T. & Kizaki, H. (2004). Streptococcus mutans strains harboring collagen-binding adhesin. *J Dent Res* **83**, 534-539.

Sattely, E. S., Fischbach, M. A. & Walsh, C. T. (2008). Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways. *Nat Prod Rep* **25**, 757-793.

Saurin, W., Hofnung, M. & Dassa, E. (1999). Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol* **48**, 22-41.

Schmitt, M. P. & Holmes, R. K. (1991). Characterization of a defective diphtheria toxin repressor (dtxR) allele and analysis of dtxR transcription in wild-type and mutant strains of Corynebacterium diphtheriae. *Infect Immun* **59**, 3903-3908.

Schmitt, M. P., Predich, M., Doukhan, L., Smith, I. & Holmes, R. K. (1995). Characterization of an iron-dependent regulatory protein (IdeR) of Mycobacterium tuberculosis as a functional homolog of the diphtheria toxin repressor (DtxR) from Corynebacterium diphtheriae. *Infect Immun* **63**, 4284-4289.

Schneider, A. & Marahiel, M. A. (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis. *Arch Microbiol* **169**, 404-410.

Schroder, I., Johnson, E. & de Vries, S. (2003). Microbial ferric iron reductases. *FEMS Microbiol Rev* **27**, 427-447.

Schubert, S., Dufke, S., Sorsa, J. & Heesemann, J. (2004). A novel integrative and conjugative element (ICE) of Escherichia coli: the putative progenitor of the Yersinia high-pathogenicity island. *Mol Microbiol* **51**, 837-848.

Schwarz-Linek, U., Werner, J. M., Pickford, A. R. & other authors (2003). Pathogenic bacteria attach to human fibronectin through a tandem beta-zipper. *Nature* **423**, 177-181.

Schwarz-Linek, U., Hook, M. & Potts, J. R. (2004). The molecular basis of fibronectin-mediated bacterial adherence to host cells. *Mol Microbiol* **52**, 631-641.

Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* **99**, 14083-14088.

Schwyn, B. & Neilands, J. B. (1987). Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* **160**, 47-56.

Scott, J. R. & Zahner, D. (2006). Pili with strong attachments: Gram-positive bacteria do it differently. *Mol Microbiol* **62**, 320-330.

Sebaihia, M., Wren, B. W., Mullany, P. & other authors (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet* **38**, 779-786.

Sebulsky, M. T. & Heinrichs, D. E. (2001). Identification and characterization of fhuD1 and fhuD2, two genes involved in iron-hydroxamate uptake in Staphylococcus aureus. *J Bacteriol* **183**, 4994-5000.

Seedorf, H., Fricke, W. F., Veith, B. & other authors (2008). The genome of Clostridium kluyveri, a strict anaerobe with unique metabolic features. *Proc Natl Acad Sci U S A* **105**, 2128-2133.

Seppala, H., Skurnik, M., Soini, H., Roberts, M. C. & Huovinen, P. (1998). A novel erythromycin resistance methylase gene (ermTR) in Streptococcus pyogenes. *Antimicrob Agents Chemother* **42**, 257-262.

Serino, L., Reimmann, C., Baur, H., Beyeler, M., Visca, P. & Haas, D. (1995). Structural genes for salicylate biosynthesis from chorismate in Pseudomonas aeruginosa. *Mol Gen Genet* **249**, 217-228.

Serino, L., Reimmann, C., Visca, P., Beyeler, M., Chiesa, V. D. & Haas, D. (1997). Biosynthesis of pyochelin and dihydroaeruginoic acid requires the iron-regulated pchDCBA operon in Pseudomonas aeruginosa. *J Bacteriol* **179**, 248-257.

Sharp, M. W., Prince, M. J. & Gibbens, J. (1995). S zooepidemicus infection and bovine mastitis. *Vet Rec* **137**, 128.

Shaw-Reid, C. A., Kelleher, N. L., Losey, H. C., Gehring, A. M., Berg, C. & Walsh, C. T. (1999). Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chem Biol* **6**, 385-400.

Shea, C. M. & McIntosh, M. A. (1991). Nucleotide sequence and genetic organization of the ferric enterobactin transport system: homology to other periplasmic binding protein-dependent systems in Escherichia coli. *Mol Microbiol* **5**, 1415-1428.

Shelburne, S. A., Davenport, M. T., Keith, D. B. & Musser, J. M. (2008a). The role of complex carbohydrate catabolism in the pathogenesis of invasive streptococci. *Trends Microbiol* **16**, 318-325.

Shelburne, S. A., 3rd, Granville, C., Tokuyama, M., Sitkiewicz, I., Patel, P. & Musser, J. M. (2005). Growth characteristics of and virulence factor production by group A Streptococcus during cultivation in human saliva. *Infect Immun* **73**, 4723-4731.

Shelburne, S. A., 3rd, Keith, D., Horstmann, N., Sumby, P., Davenport, M. T., Graviss, E. A., Brennan, R. G. & Musser, J. M. (2008b). A direct link between

carbohydrate utilization and virulence in the major human pathogen group A *Streptococcus*. *Proc Natl Acad Sci U S A* **105**, 1698-1703.

Sheoran, A. S., Sponseller, B. T., Holmes, M. A. & Timoney, J. F. (1997). Serum and mucosal antibody isotype responses to M-like protein (SeM) of Streptococcus equi in convalescent and vaccinated horses. *Vet Immunol Immunopathol* **59**, 239-251.

Sheoran, A. S., Timoney, J. F., Holmes, M. A., Karzenski, S. S. & Crisman, M. V. (2000). Immunoglobulin isotypes in sera and nasal mucosal secretions and their neonatal transfer and distribution in horses. *Am J Vet Res* **61**, 1099-1105.

Sheoran, A. S., Artiushin, S. & Timoney, J. F. (2002). Nasal mucosal immunogenicity for the horse of a SeM peptide of Streptococcus equi genetically coupled to cholera toxin. *Vaccine* **20**, 1653-1659.

Sherman, J. M. (1937). The Streptococci. *Bacteriol Rev* **1**, 3-97.

Silakowski, B., Schairer, H. U., Ehret, H. & other authors (1999). New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1. *J Biol Chem* **274**, 37391-37399.

Silakowski, B., Kunze, B., Nordsiek, G., Blocker, H., Hofle, G. & Muller, R. (2000). The myxochelin iron transport regulon of the myxobacterium Stigmatella aurantiaca Sg a15. *Eur J Biochem* **267**, 6476-6485.

Sitkiewicz, I., Nagiec, M. J., Sumby, P., Butler, S. D., Cywes-Bentley, C. & Musser, J. M. (2006). Emergence of a bacterial clone with enhanced virulence by acquisition of a phage encoding a secreted phospholipase A2. *Proc Natl Acad Sci U S A* **103**, 16009-16014.

Slater, J. D. (2003). Strangles, bastard strangles, vives and glanders: archaeological relics in a genomic age. *Equine Vet J* **35**, 118-120.

Smith, K. C., Blunden, A. S., Whitwell, K. E., Dunn, K. A. & Wales, A. D. (2003). A survey of equine abortion, stillbirth and neonatal death in the UK from 1988 to 1997. *Equine Vet J* **35**, 496-501.

Soderberg, J. J. & von Pawel-Rammingen, U. (2008). The streptococcal protease IdeS modulates bacterial IgGFc binding and generates 1/2Fc fragments with the ability to prime polymorphonuclear leucocytes. *Mol Immunol* **45**, 3347-3353.

Soedarmanto, I., Pasaribu, F. H., Wibawan, I. W. & Lammler, C. (1996). Identification and molecular characterization of serological group C streptococci isolated from diseased pigs and monkeys in Indonesia. *J Clin Microbiol* **34**, 2201-2204.

Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X. R., Krogfelt, K. A., Struve, C., Schembri, M. A. & Hasty, D. L. (1998). Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. *Proc Natl Acad Sci U S A* **95**, 8922-8926.

Sorek, R., Kunin, V. & Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**, 181-186.

Srivastava, S. K. & Barnum, D. A. (1983). Adherence of Streptococcus equi on tongue, cheek and nasal epithelial cells of ponies. *Vet Microbiol* **8**, 493-504.

Stableforth, A. W. (1959). Streptococcal diseases. In *Infectious diseases of animals Diseases due to bacteria*, pp. 589-650. Edited by A. W. a. G. Stableforth, I. A. . New York: Academic press.

Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493-505.

Starr, C. R. & Engleberg, N. C. (2006). Role of hyaluronidase in subcutaneous spread and growth of group A streptococcus. *Infect Immun* **74**, 40-48.

Steichen, C., Chen, P., Kearney, J. F. & Turnbough, C. L., Jr. (2003). Identification of the immunodominant protein and other proteins of the Bacillus anthracis exosporium. *J Bacteriol* **185**, 1903-1910.

Stevenson, R. G. (1974). Streptococcus zooepidemicus infection in sheep. *Can J Comp Med* **38**, 243-250.

Sundaramoorthy, R., Fyfe, P. K. & Hunter, W. N. (2008). Structure of Staphylococcus aureus EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein. *J Mol Biol* **383**, 603-614.

Sweeney, C. R., Divers, T. J. & Benson, C. E. (1985). Anaerobic bacteria in 21 horses with pleuropneumonia. *J Am Vet Med Assoc* **187**, 721-724.

Sweeney, C. R., Whitlock, R. H., Meirs, D. A., Whitehead, S. C. & Barningham, S. O. (1987). Complications associated with Streptococcus equi infection on a horse farm. *J Am Vet Med Assoc* **191**, 1446-1448.

Sweeney, C. R., Holcombe, S. J., Barningham, S. C. & Beech, J. (1991). Aerobic and anaerobic bacterial isolates from horses with pneumonia or pleuropneumonia and antimicrobial susceptibility patterns of the aerobes. *J Am Vet Med Assoc* **198**, 839-842.

Sweeney, C. R., Timoney, J. F., Newton, J. R. & Hines, M. T. (2005). Streptococcus equi infections in horses: guidelines for treatment, control, and prevention of strangles. *J Vet Intern Med* **19**, 123-134.

Sylvestre, P., Couture-Tosi, E. & Mock, M. (2002). A collagen-like surface glycoprotein is a structural component of the Bacillus anthracis exosporium. *Mol Microbiol* **45**, 169-178.

Tamura, G. S., Hull, J. R., Oberg, M. D. & Castner, D. G. (2006). High-affinity interaction between fibronectin and the group B streptococcal C5a peptidase is unaffected by a naturally occurring four-amino-acid deletion that eliminates peptidase activity. *Infect Immun* **74**, 5739-5746.

Tanabe, T., Nakao, H., Kuroda, T., Tsuchiya, T. & Yamamoto, S. (2006). Involvement of the Vibrio parahaemolyticus pvsC gene in export of the siderophore vibrioferrin. *Microbiol Immunol* **50**, 871-876.

Taron, C. H., Benner, J. S., Hornstra, L. J. & Guthrie, E. P. (1995). A novel beta-galactosidase gene isolated from the bacterium Xanthomonas manihotis exhibits strong homology to several eukaryotic beta-galactosidases. *Glycobiology* **5**, 603-610.

Tekaia, F., Gordon, S. V., Garnier, T., Brosch, R., Barrell, B. G. & Cole, S. T. (1999). Analysis of the proteome of Mycobacterium tuberculosis in silico. *Tuber Lung Dis* **79**, 329-342.

Telford, J. L., Barocchi, M. A., Margarit, I., Rappuoli, R. & Grandi, G. (2006). Pili in gram-positive pathogens. *Nat Rev Microbiol* **4**, 509-519.

Tettelin, H., Nelson, K. E., Paulsen, I. T. & other authors (2001). Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science* **293**, 498-506.

Theisen, M., Borre, M., Mathiesen, M. J., Mikkelsen, B., Lebech, A. M. & Hansen, K. (1995). Evolution of the Borrelia burgdorferi outer surface protein OspC. *J Bacteriol* **177**, 3036-3044.

Thompson, B. M. & Stewart, G. C. (2008). Targeting of the BclA and BclB proteins to the Bacillus anthracis spore surface. *Mol Microbiol* **70**, 421-434.

Tillier, E. R. & Collins, R. A. (2000). Genome rearrangement by replication-directed translocation. *Nat Genet* **26**, 195-197.

Timoney, J. F. & Eggers, D. (1985). Serum bactericidal responses to Streptococcus equi of horses following infection or vaccination. *Equine Vet J* **17**, 306-310.

Timoney, J. F. & Trachman, J. (1985). Immunologically reactive proteins of Streptococcus equi. *Infect Immun* **48**, 29-34.

Timoney, J. F. (1993). Strangles. *Vet Clin North Am Equine Pract* **9**, 365-374.

Timoney, J. F. & Mukhtar, M. M. (1993). The protective M proteins of the equine group C streptococci. *Vet Microbiol* **37**, 389-395.

Timoney, J. F., Walker, J., Zhou, M. & Ding, J. (1995). Cloning and sequence analysis of a protective M-like protein gene from Streptococcus equi subsp. zooepidemicus. *Infect Immun* **63**, 1440-1445.

Timoney, J. F. (1997). Serology of *Streptococcus equi* infection: uses and limitations. In *Equine Disease Quarterly*, pp. 2-3. Kentucky: University of Kentucky.

Timoney, J. F., Anzai, T. & Blair, M. (1997a). Clonal invasion of the equine respiratory tract by Streptococcus zooepidemicus. *Adv Exp Med Biol* **418**, 611-613.

Timoney, J. F. & Artiushin, S. C. (1997). Detection of Streptococcus equi in equine nasal swabs and washes by DNA amplification. *Vet Rec* **141**, 446-447.

Timoney, J. F., Artiushin, S. C. & Boschwitz, J. S. (1997b). Comparison of the sequences and functions of Streptococcus equi M-like proteins SeM and SzPSe. *Infect Immun* **65**, 3600-3605.

Timoney, J. F., Sheoran, A. & Artiushin, S. (1998). Detection of strangles carriers. *Vet Rec* **142**, 648.

Timoney, J. F. (2004). The pathogenic equine streptococci. *Vet Res* **35**, 397-409.

Timoney, J. F., Qin, A., Muthupalani, S. & Artiushin, S. (2007). Vaccine potential of novel surface exposed and secreted proteins of Streptococcus equi. *Vaccine* **25**, 5583-5590.

Timoney, J. F. & Kumar, P. (2008). Early pathogenesis of equine Streptococcus equi infection (strangles). *Equine Vet J* **40**, 637-642.

Timoney, J. F., Yang, J., Liu, J. & Merant, C. (2008). IdeE reduces the bactericidal activity of equine neutrophils for Streptococcus equi. *Vet Immunol Immunopathol* **122**, 76-82.

Timoney, J. F., Artuishin, S., Wang, J. (1998). Regions of the protective M-like protein of *Streptococcus equi* recognised by serum and mucosal antibodies of convalescent horses.

In *Equine Infectious Diseases VIII.* Edited by U. Wernery, Wade, J. F., Mumford, J. A., Kaaden, O. R. Dubai, United Arab Emirates: R & W Publications, Newmarket.

Timoney, J. F., Gillespie, J. H., Scott, F. W., Barlough, J. E. (1988). Hagan and Bruner's Microbiology and Infectious Diseases of Domestic Animals: Cornell University Press.

Tiwari, R., Qin, A., Artiushin, S. & Timoney, J. F. (2007). Se18.9, an anti-phagocytic factor H binding protein of *Streptococcus equi. Vet Microbiol* **121**, 105-115.

Tolmasky, M. E., Actis, L. A. & Crosa, J. H. (1993). A single amino acid change in AngR, a protein encoded by pJM1-like virulence plasmids, results in hyperproduction of anguibactin. *Infect Immun* **61**, 3228-3233.

Tolmasky, M. E., Wertheimer, A. M., Actis, L. A. & Crosa, J. H. (1994). Characterization of the Vibrio anguillarum fur gene: role in regulation of expression of the FatA outer membrane protein and catechols. *J Bacteriol* **176**, 213-220.

Ton-That, H. & Schneewind, O. (2003). Assembly of pili on the surface of Corynebacterium diphtheriae. *Mol Microbiol* **50**, 1429-1438.

Ton-That, H., Marraffini, L. A. & Schneewind, O. (2004). Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae. *Mol Microbiol* **53**, 251-261.

Toussaint, A. & Merlin, C. (2002). Mobile elements as a combination of functional modules. *Plasmid* **47**, 26-35.

Tseng, C. F., Burger, A., Mislin, G. L., Schalk, I. J., Yu, S. S., Chan, S. I. & Abdallah, M. A. (2006). Bacterial siderophores: the solution stoichiometry and coordination of the Fe(III) complexes of pyochelin and related compounds. *J Biol Inorg Chem* **11**, 419-432.

Turlan, C. & Chandler, M. (1995). IS1-mediated intramolecular rearrangements: formation of excised transposon circles and replicative deletions. *EMBO J* **14**, 5410-5421.

Turner, K. M., Hanage, W. P., Fraser, C., Connor, T. R. & Spratt, B. G. (2007). Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC Microbiol* 7, 30.

Ulrich, R. G., Olson, M. A. & Bavari, S. (1998). Development of engineered vaccines effective against structurally related bacterial superantigens. *Vaccine* **16**, 1857-1864.

Valentin-Weigand, P., Traore, M. Y., Blobel, H. & Chhatwal, G. S. (1990). Role of alpha 2-macroglobulin in phagocytosis of group A and C streptococci. *FEMS Microbiol Lett* **58**, 321-324.

van der Heide, T. & Poolman, B. (2002). ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep* **3**, 938-943.

van der Woude, M. W. & Baumler, A. J. (2004). Phase and antigenic variation in bacteria. *Clin Microbiol Rev* **17**, 581-611, table of contents.

van Loo, I. H., Heuvelman, K. J., King, A. J. & Mooi, F. R. (2002). Multilocus sequence typing of Bordetella pertussis based on surface protein genes. *J Clin Microbiol* **40**, 1994-2001.

Varhimo, E., Savijoki, K., Jalava, J., Kuipers, O. P. & Varmanen, P. (2007). Identification of a novel streptococcal gene cassette mediating SOS mutagenesis in Streptococcus uberis. *J Bacteriol* **189**, 5210-5222.

Vartivarian, S. E. & Cowart, R. E. (1999). Extracellular iron reductases: identification of a new class of enzymes by siderophore-producing microorganisms. *Arch Biochem Biophys* **364**, 75-82.

Venturi, V., Ottevanger, C., Bracke, M. & Weisbeek, P. (1995). Iron regulation of siderophore biosynthesis and transport in Pseudomonas putida WCS358: involvement of a transcriptional activator and of the Fur protein. *Mol Microbiol* **15**, 1081-1093.

Vickerman, M. M., Iobst, S., Jesionowski, A. M. & Gill, S. R. (2007). Genome-wide transcriptional changes in Streptococcus gordonii in response to competence signaling peptide. *J Bacteriol* **189**, 7799-7807.

Visca, P., Colotti, G., Serino, L., Verzili, D., Orsi, N. & Chiancone, E. (1992). Metal regulation of siderophore synthesis in Pseudomonas aeruginosa and functional effects of siderophore-metal complexes. *Appl Environ Microbiol* **58**, 2886-2893.

von Dohren, H., Keller, U., Vater, J. & Zocher, R. (1997). Multifunctional Peptide Synthetases. *Chem Rev* **97**, 2675-2706.

von Pawel-Rammingen, U., Johansson, B. P. & Bjorck, L. (2002). IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *EMBO J* **21**, 1607-1615.

Wagner, B., Greiser-Wilke, I., Wege, A. K., Radbruch, A. & Leibold, W. (2002). Evolution of the six horse IGHG genes and corresponding immunoglobulin gamma heavy chains. *Immunogenetics* **54**, 353-364.

Wagner, B., Miller, D. C., Lear, T. L. & Antczak, D. F. (2004). The complete map of the Ig heavy chain constant gene region reveals evidence for seven IgG isotypes and for IgD in the horse. *J Immunol* **173**, 3230-3242.

Walker, J. A. & Timoney, J. F. (1998). Molecular basis of variation in protective SzP proteins of Streptococcus zooepidemicus. *Am J Vet Res* **59**, 1129-1133.

Walker, J. A. & Timoney, J. F. (2002). Construction of a stable non-mucoid deletion mutant of the Streptococcus equi Pinnacle vaccine strain. *Vet Microbiol* **89**, 311-321.

Waller, A., Flock, M., Smith, K. & other authors (2007). Vaccination of horses against strangles using recombinant antigens from Streptococcus equi. *Vaccine* **25**, 3629-3635.

Waller, A. S. (2007). Bacterial adherence to host tissues: a very sticky problem. *Vet J* **173**, 478-479.

Waller, A. S. & Jolley, K. A. (2007). Getting a grip on strangles: recent progress towards improved diagnostics and vaccines. *Vet J* **173**, 492-501.

Waller, L. N., Stump, M. J., Fox, K. F., Harley, W. M., Fox, A., Stewart, G. C. & Shahgholi, M. (2005). Identification of a second collagen-like glycoprotein produced by Bacillus anthracis and demonstration of associated spore-specific sugars. *J Bacteriol* **187**, 4592-4597.

Wandersman, C. & Delepelaire, P. (2004). Bacterial iron sources: from siderophores to hemophores. *Annu Rev Microbiol* **58**, 611-647.

Wang, H., Roberts, A. P., Lyras, D., Rood, J. I., Wilks, M. & Mullany, P. (2000). Characterization of the ends and target sites of the novel conjugative transposon Tn5397 from Clostridium difficile: excision and circularization is mediated by the large resolvase, TndX. *J Bacteriol* **182**, 3775-3783.

Ward, P. N., Holden, M. T., Leigh, J. A. & other authors (2009). Evidence for niche adaptation in the genome of the bovine pathogen Streptococcus uberis. *BMC Genomics* **10**, 54.

Watanabe, K., Rude, M. A., Walsh, C. T. & Khosla, C. (2003). Engineered biosynthesis of an ansamycin polyketide precursor in Escherichia coli. *Proc Natl Acad Sci U S A* **100**, 9774-9778.

Watanabe, K., Hotta, K., Praseuth, A. P., Koketsu, K., Migita, A., Boddy, C. N., Wang, C. C., Oguri, H. & Oikawa, H. (2006). Total biosynthesis of antitumor nonribosomal peptides in Escherichia coli. *Nat Chem Biol* **2**, 423-428.

Way, S. S. & Wilson, C. B. (2005). The Mycobacterium tuberculosis ESAT-6 homologue in Listeria monocytogenes is dispensable for growth in vitro and in vivo. *Infect Immun* **73**, 6151-6153.

Webb, K., Jolley, K. A., Mitchell, Z., Robinson, C., Newton, J. R., Maiden, M. C. & Waller, A. (2008). Development of an unambiguous and discriminatory multilocus

sequence typing scheme for the Streptococcus zooepidemicus group. *Microbiology* **154**, 3016-3024.

Weber, G., Schorgendorfer, K., Schneider-Scherzer, E. & Leitner, E. (1994). The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet* **26**, 120-125.

Wei, J., Goldberg, M. B., Burland, V. & other authors (2003). Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T. *Infect Immun* **71**, 2775-2786.

Welsh, R. D. (1984). The significance of *Streptococcus zooepidemicus* in the horse. *Equine Practice* **6**, 1129-1133.

Wessels, M. R., Moses, A. E., Goldberg, J. B. & DiCesare, T. J. (1991). Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci. *Proc Natl Acad Sci U S A* **88**, 8317-8321.

Wessels, M. R. (2000). Gram-positive pathogens, pp. 34-42. Edited by V. A. N. Fischetti, R. P. Ferretti, J. J. Portnoy, D. A. Rood, J. I. . Washington DC: ASM Press.

Wexler, D. E., Chenoweth, D. E. & Cleary, P. P. (1985). Mechanism of action of the group A streptococcal C5a inactivator. *Proc Natl Acad Sci U S A* **82**, 8144-8148.

Whittam, T. S. & Bumbaugh, A. C. (2002). Inferences from whole-genome sequences of bacterial pathogens. *Curr Opin Genet Dev* **12**, 719-725.

Wibawan, I. W., Pasaribu, F. H., Utama, I. H., Abdulmawjood, A. & Lammler, C. (1999). The role of hyaluronic acid capsular material of Streptococcus equi subsp. zooepidemicus in mediating adherence to HeLa cells and in resisting phagocytosis. *Res Vet Sci* **67**, 131-135.

Widders, P. R., Warner, S. & Huntington, P. J. (1995). Immunisation of mares to control endometritis caused by Streptococcus zooepidemicus. *Res Vet Sci* **58**, 75-81.

Wisniewski-Dye, F. & Vial, L. (2008). Phase and antigenic variation mediated by genome modifications. *Antonie Van Leeuwenhoek* **94**, 493-515.

Woo, P. C., To, A. P., Tse, H., Lau, S. K. & Yuen, K. Y. (2003). Clinical and molecular epidemiology of erythromycin-resistant beta-hemolytic lancefield group G streptococci causing bacteremia. *J Clin Microbiol* **41**, 5188-5191.

Wood, J. L., Burrell, M. H., Roberts, C. A., Chanter, N. & Shaw, Y. (1993). Streptococci and Pasteurella spp. associated with disease of the equine lower respiratory tract. *Equine Vet J* **25**, 314-318.

Wood, J. L., Newton, J. R., Chanter, N. & Mumford, J. A. (2005a). Association between respiratory disease and bacterial and viral infections in British racehorses. *J Clin Microbiol* **43**, 120-126.

Wood, J. L., Newton, J. R., Chanter, N. & Mumford, J. A. (2005b). Inflammatory airway disease, nasal discharge and respiratory infections in young British racehorses. *Equine Vet J* **37**, 236-242.

Wood, J. L. N. (1999). An epidemiological investigation of respiratory disease in racehorses. In *Life Sciences*, pp. 281: Open University.

Wood, J. L. N., Chanter, N. (1994). Can washing help keep the lungs clean. *Equine Vet Educ* **6**, 220-222.

Woolcock, J. B. (1975). Epidemiology of equine streptococci. *Res Vet Sci* **18**, 113-114.

Wooldridge, K. G. & Williams, P. H. (1993). Iron uptake mechanisms of pathogenic bacteria. *FEMS Microbiol Rev* **12**, 325-348.

Wyckoff, E. E., Stoebner, J. A., Reed, K. E. & Payne, S. M. (1997). Cloning of a Vibrio cholerae vibriobactin gene cluster: identification of genes required for early steps in siderophore biosynthesis. *J Bacteriol* **179**, 7055-7062.

Wyckoff, E. E., Valle, A. M., Smith, S. L. & Payne, S. M. (1999). A multifunctional ATP-binding cassette transporter system from Vibrio cholerae transports vibriobactin and enterobactin. *J Bacteriol* **181**, 7588-7596.

Xu, P., Alves, J. M., Kitten, T. & other authors (2007). Genome of the opportunistic pathogen Streptococcus sanguinis. *J Bacteriol* **189**, 3166-3175.

Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M. & Lukomski, S. (2002). Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem* **277**, 27312-27318.

Xu, Y., Liang, X., Chen, Y., Koehler, T. M. & Hook, M. (2004). Identification and biochemical characterization of two novel collagen binding MSCRAMMs of Bacillus anthracis. *J Biol Chem* **279**, 51760-51768.

Yarwood, J. M., McCormick, J. K., Paustian, M. L., Orwin, P. M., Kapur, V. & Schlievert, P. M. (2002). Characterization and expression analysis of Staphylococcus aureus pathogenicity island 3. Implications for the evolution of staphylococcal pathogenicity islands. *J Biol Chem* **277**, 13138-13147.

Yoshikawa, H., Yasu, T., Ueki, H., Oyamada, T., Oishi, H., Anzai, T., Oikawa, M. & Yoshikawa, T. (2003). Pneumonia in horses induced by intrapulmonary inoculation of Streptococcus equi subsp. zooepidemicus. *J Vet Med Sci* **65**, 787-792.

Youard, Z. A., Mislin, G. L., Majcherczyk, P. A., Schalk, I. J. & Reimmann, C. (2007). Pseudomonas fluorescens CHA0 produces enantio-pyochelin, the optical antipode of the Pseudomonas aeruginosa siderophore pyochelin. *J Biol Chem* **282**, 35546-35553.

Yu, S., Fiss, E. & Jacobs, W. R., Jr. (1998). Analysis of the exochelin locus in Mycobacterium smegmatis: biosynthesis genes have homology with genes of the peptide synthetase family. *J Bacteriol* **180**, 4676-4685.

Zhang, J. R. & Norris, S. J. (1998). Genetic variation of the Borrelia burgdorferi gene vlsE involves cassette-specific, segmental gene conversion. *Infect Immun* **66**, 3698-3704.

Zhang, Q. & Wise, K. S. (1996). Molecular basis of size and antigenic variation of a Mycoplasma hominis adhesin encoded by divergent vaa genes. *Infect Immun* **64**, 2737-2744.

Zhu, H., Liu, M. & Lei, B. (2008). The surface protein Shr of Streptococcus pyogenes binds heme and transfers it to the streptococcal heme-binding protein Shp. *BMC Microbiol* **8**, 15.

Zhu, W., Arceneaux, J. E., Beggs, M. L., Byers, B. R., Eisenach, K. D. & Lundrigan, M. D. (1998). Exochelin genes in Mycobacterium smegmatis: identification of an ABC transporter and two non-ribosomal peptide synthetase genes. *Mol Microbiol* **29**, 629-639.

Zinkernagel, A. S., Timmer, A. M., Pence, M. A. & other authors (2008). The IL-8 protease SpyCEP/ScpC of group A Streptococcus promotes resistance to neutrophil killing. *Cell Host Microbe* **4**, 170-178.

Appendix

Table A.1 Identity, source and mitogenicity of *S. equi* and *S. zooepidemicus* isolates studied for the prevalence of selected differences between the *Se*4047 and *Sz*H70 genomes

Species	ST	Strain ID	Disease/ Source	Animal	Year	Stimulation index
Sz	19	B27 3726.1	Uterine infection/abortion	Horse	2007	1.5
Sz	101	3604	Uterine infection/abortion	Horse	2006	1.2
Sz	101	B24 2382	Uterine infection/abortion	Horse	2004	1.6
Sz	101	B24 4996	Uterine infection/abortion	Horse	2004	1.7
Sz	142	3727	Uterine infection/abortion	Horse	2007	1.1
Sz	56	B23 3322	Uterine infection/abortion	Horse	2003	1.7
Se	179	3154	Strangles	Horse	2004	2.8
Se	179	8229	Strangles	Dog	2004	3.3
Se	179	B23 7325	Strangles	Horse	2003	2.5
Se	179	4047	Strangles	Horse	1990	4.2
Se	179	SA	Strangles	Horse	1999	2.9
Se	179	303	Strangles	Horse	1999	4
Se	179	7060	Strangles	Horse	2003	2.9
Se	179	1931	Strangles	Horse	2004	4.3
Se	179	1351	Strangles	Horse	2004	3.1
Se	179	1350	Strangles	Horse	2003	3.3
Se	179	7344	Strangles	Horse	2003	2.8
Se	179	7140	Strangles	Horse	2003	2.8
Se	179	7171	Strangles	Horse	2003	3.5
Se	179	7364	Strangles	Horse	2003	3.1
Se	179	7326	Strangles	Horse	2003	2.3
Se	179	3155	Strangles	Horse	2004	3.6
Se	179	3156	Strangles	Horse	2004	3.2
Se	179	181063	Strangles	Horse	1999	3
Se	179	JKS 044	Strangles carrier	Horse	2006	2.6
Se	179	JKS 043	Strangles	Horse	2006	3.4
Se	179	JKS 063	Strangles	Horse	2006	3
Se	179	JKS 225	Strangles	Horse	2006	3.2
Se	179	CF32	Strangles	Horse	1981	2.6
Se	179	JKS 559	Strangles	Horse	2007	2.2
Se	151	7329	Strangles	Horse	2006	2.7
Se	151	Mo	Strangles brain abscess	Horse	2008	2.7
Sz	49	B25 4861.1	Wound infection	Horse	2005	0.9
Sz	49	B23 6378	Uterine infection/abortion	Horse	2003	0.8
Sz	49	B27 1175.2	Uterine infection/abortion	Horse	2007	1.1
Sz	49	B26 8276	Upper respiratory	Horse	2006	1.3
Sz	137	628	Uterine infection/abortion	Horse	2007	1.0
Sz	149	6619	Uterine infection/abortion	Horse	2003	1.4
Sz	6	S13	Skin swab	Dog	2002	1.7
Sz	133	2258	Foot swab	Horse	2007	1.6

Table A.1 continued

Species	ST	Strain ID	Disease/ Source	Animal	Year	Stimulation index
Sz	10	BHS32	Lower respiratory	Dog	2001	1.0
Sz	10	BHS41	Lower respiratory	Dog	2001	1
Sz	10	B26 6443	Wound infection	Horse	2006	1.3
Sz	22	BHS53	Lower respiratory	Dog	2001	1.1
Sz	47	BHS 28	Lower respiratory	Dog	2000	1.2
Sz	55	B23 7163	Uterine infection/abortion	Horse	2003	2.1
Sz	15	B23 6907	Uterine infection/abortion	Horse	2003	1.1
Sz	15	B27 2806.2	Uterine infection/abortion	Horse	2007	1.0
Sz	91	BHS 37	Lower respiratory	Dog	2001	1.1
Sz	110	4847	Upper respiratory	Horse	1996	1.5
Sz	119	4904	Upper respiratory	Horse	1996	1.1
Sz	119	5382	Upper respiratory	Horse	1996	1.0
Sz	45	605851	Upper respiratory	Horse	1996	1.2
Sz	45	5845	Upper respiratory	Horse	1996	1.5
Sz	45	D56	Lower respiratory	Horse	2000	0.7
Sz	147	6875	Uterine infection/abortion	Horse	2006	1.4
Sz	51	B23 6074	Uterine infection/abortion	Horse	2003	1.7
Sz	143	3512	Uterine infection/abortion	Horse	2007	1.3
Sz	128	8718.2	Uterine infection/abortion	Horse	2006	1.9
Sz	111	5831	Upper respiratory	Horse	1996	1.8
Sz	126	5079	Uterine infection/abortion	Horse	2006	2.4
Sz	131	3543	Foot swab	Horse	2006	0.9
Sz	16	B24 7132	Upper respiratory	Horse	2004	1.1
Sz	140	B27 3530.1	Wound infection	Horse	2007	1.3
Sz	134	2567.1	Udder swab	Horse	2007	1.4
Sz	57	JKS 115	Upper respiratory	Horse	2006	2.4
Sz	1	H70	Upper respiratory	Horse	2000	1.4
Sz	1	B23 7271	Uterine infection/abortion	Horse	2003	1.5
Sz	1	B27 0479	Uterine infection/abortion	Horse	2007	1.3
Sz	71	60 5632	Upper respiratory	Horse	1996	1.0
Sz	71	60 5013	Upper respiratory	Horse	1996	1.3
Sz	104	8278	Upper respiratory	Horse	2006	1.2
Sz	104	8275	Upper respiratory	Horse	2006	1.2
Sz	108	5938	Upper respiratory	Horse	1996	1.3
Sz	108	4853	Upper respiratory	Horse	1996	0.9
Sz	123	BHS 5	Lower respiratory	Dog	1999	6.7
Sz	127	6360	Wound infection	Horse	2006	5.0
Sz	141	482	Uterine infection/abortion	Horse	2007	5.8
Sz	7	B26 8900	Wound infection	Horse	2006	5.5
Sz	7	B27 1185	Uterine infection/abortion	Horse	2007	3.5
Sz	7	60 2333	Upper respiratory	Horse	1996	4.9
Sz	48	B25 4433	Lymph node abscess	Horse	2005	3.3
Sz	70	B26 8310	Upper respiratory	Horse	2006	3.3
Sz	70	B26 8277	Upper respiratory	Horse	2006	2.6
Sz	5	B24 7159	Lymph node abscess	Horse	2004	5.4
Sz	5	B24 7043	Lymph node abscess	Horse	2004	2.9
Sz	53	B24 7102.2	Lymph node abscess	Horse	2004	7.1

Table A.1 continued

Species	ST	Strain ID	Disease/ Source	Animal	Year	Stimulation index
Sz	8	H8	Lower respiratory	Horse	2000	9.6
Sz	8	D40	Lower respiratory	Horse	2000	5.3
Sz	8	D2a	Lower respiratory	Horse	2000	5.1
Sz	46	B25 0061	Lymph node abscess	Horse	2005	3.6
Sz	113	4885	Upper respiratory	Horse	1996	6.5
Sz	113	5617	Upper respiratory	Horse	1996	5.3
Sz	96	B25 4763	Lymph node abscess	Horse	2005	1.0
Sz	20	B25 4042.1	NA	Horse	2005	1.7
Sz	118	3050	Nasal fluid	Horse	2007	1.5
Sz	118	4901	Upper respiratory	Horse	1996	1.4
Sz	118	2329	Upper respiratory	Horse	1996	2.0
Sz	139	972395	Upper respiratory	Horse	1997	0.8
Sz	146	8250	Wound infection	Horse	2006	1.1
Sz	58	B26 Tansey	Nephritis	Horse	2006	1.3
Sz	61	B26 8269.1	Wound infection	Horse	2006	1.0
Sz	61	B26 0225	Uterine infection/abortion	Horse	2006	1.1
Sz	94	B23 6170	Uterine infection/abortion	Horse	2003	1.7
Sz	26	B27 2596	Uterine infection/abortion	Horse	2007	1.3
Sz	26	B26 6993	Uterine infection/abortion	Horse	2006	1.6
Sz	124	5808	Wound infection	Horse	2003	0.7
Sz	135	2265	Intra-abdomen pus	Horse	2007	1.8
Sz	39	B26 6334	Wound infection	Horse	2006	1.0
Sz	39	JKS 241	Upper respiratory	Horse	2007	1.3
Sz	106	605849	Upper respiratory	Horse	1996	8.0
Sz	106	5936	Upper respiratory	Horse	1996	3.2
Sz	106	8306	Upper respiratory	Horse	2006	1.5
Sz	107	8307	Upper respiratory	Horse	2006	0.7
Sz	9	H54	Upper respiratory	Horse	2000	1.9
Sz	120	4897	Upper respiratory	Horse	1996	3.5
Sz	63	B26 2253	Uterine infection/abortion	Horse	2006	1.4
Sz	63	B26 3540	Foot swab	Horse	2006	1.2
Sz	54	B24 7155	Upper respiratory	Horse	2004	1.6
Sz	130	7101	Wound infection	Horse	2006	1.6
Sz	50	B25 5405	Lymph node abscess	Horse	2005	1.5
Sz	138	917	Uterine infection/abortion	Horse	2007	1.0
Sz	82	B26 6458	Uterine infection/abortion	Horse	2006	1.6
Sz	178	2958	Uterine infection/abortion	Horse	2003	1.6
Sz	109	5820	Upper respiratory	Horse	1996	1.3
Sz	150	8311	Upper respiratory	Horse	2006	1.0
Sz	150	8308	Upper respiratory	Horse	2006	1.8
Sz	4	B27 4388	Uterine infection/abortion	Horse	2007	1.3
Sz	4	B24 4389	Uterine infection/abortion	Horse	2004	1.4
Sz	132	2853	Lymph node abscess	Horse	2007	0.8
Sz	132	1913	Uterine infection/abortion	Horse	2007	1.0
Sz	103	8300	Upper respiratory	Horse	2006	1.0
Sz	103	8297	Upper respiratory	Horse	2006	1.2
Sz	116	8293	Upper respiratory	Horse	1996	1.1
Sz	112	4866	Upper respiratory	Horse	1996	1.0
Sz	112	5951	Upper respiratory	Horse	1996	1.2
Sz	122	4871	Upper respiratory	Horse	1996	1.4
Sz	93	B23 7166	Uterine infection/abortion	Horse	2003	1.0
Sz	121	5185	Upper respiratory	Horse	2006	1.9

Table A.1 continued

Species	ST	Strain ID	Disease/ Source	Animal	Year	Stimulation index
Sz	3	D33	Lower respiratory	Horse	2000	1.8
Sz	92	B23 4314B	NA	Horse	2003	2.0
Sz	125	7157	Upper respiratory	Horse	2004	1.6
Sz	136	3101	Uterine infection/abortion	Horse	2007	0.8
Sz	97	B26 5081.1	Uterine infection/abortion	Horse	2006	1.0
Sz	97	B26 8570	Uterine infection/abortion	Horse	2006	1.2
Sz	97	B26 8571	Uterine infection/abortion	Horse	2006	1.6
Sz	97	B27 2742.2	Uterine infection/abortion	Horse	2007	1.1
Sz	97	B27 7714	Lymph node abscess	Horse	2007	1.5
Sz	97	B27 0624.2	Uterine infection/abortion	Horse	2007	1.1
Sz	97	B27 0624.1	Uterine infection/abortion	Horse	2007	1.6
Sz	97	B26 8575	Uterine infection/abortion	Horse	2006	1.4
Sz	99	B27 0739	Uterine infection/abortion	Horse	2007	1.4
Sz	117	5623	Upper respiratory	Horse	1996	1.8
Sz	148	4875	Upper respiratory	Horse	1996	1.3
Sz	2	D14a	Lower respiratory	Horse	2000	2.4
Sz	12	B25 0590	Wound infection	Horse	2005	0.9
Sz	13	B26 0863	Uterine infection/abortion	Horse	2006	1.7
Sz	18	1770	Fatal haemorrhagic pneumoniae	Dog	2008	0.9
Sz	18	1727	Fatal haemorrhagic pneumoniae	Dog	2008	1.0
Sz	95	B24 7156	Upper respiratory	Horse	2004	1.1
Sz	98	B26 8337	Keratitis	Horse	2006	1.1
Sz	98	B27 2247	Uterine infection/abortion	Horse	2007	0.8
Sz	100	B26 3593	Keratitis	Horse	2006	1.5
Sz	102	8299	Upper respiratory	Horse	2006	1.8
Sz	115	4893	Upper respiratory	Horse	1996	2.0
Sz	144	2410	Upper respiratory	Horse	1997	3.0

Data are ordered as in Figure 2.6. Horizontal line breaks separate phylogenetic clusters of *S. zooepidemicus* (*Sz*) and *S. equi* (*Se*) isolates identified by ClonalFrame analysis of MLST alleles, with the exception of the last 12 isolates, which represent 10 unrelated sequence types (STs). NA Information not available. Details of all of these isolates are also available on the online MLST database: http://pubmlst.org/szooepidemicus/. Accessed 23.02.09). Stimulation index data indicates the level of mitogenic activity towards equine PBMCs detected in supernatant samples following overnight culture of *S. zooepidemicus* or *S. equi* isolates (provided by Romain Paillot, AHT).

Table A.2A Novel *Se*4047 DNA loci

No.	Element	Notes	CDS / region	Function	Total no. novel CDS	% GC
1	putative genomic island		SEQ0047*	putative ABC transporter ATP-binding / permease protein	4	29.12
			SEQ0048*	ThiF family protein (pseudogene)		
			SEQ0050*	insulinase family metallopeptidase		
			SEQ0051*	putative regulatory protein		
2			SEQ0100	transposase (pseudogene)	2	30.11
			SEQ0101*	putative exported protein		
3	Prophage φSeq1	insertion in *Se*4047	SEQ0133	integrase	65	40.22
			SEQ0134	hypothetical phage protein		
			SEQ0135	phage repressor protein		
			SEQ0136	putative phage repressor		
			SEQ0137	hypothetical phage protein		
			SEQ0138	hypothetical phage protein		
			SEQ0139	putative DNA-binding phage protein		
			SEQ0140	hypothetical phage protein		
			SEQ0141	hypothetical phage protein		
			SEQ0142	putative DNA-binding phage protein		
			SEQ0143	hypothetical phage protein		
			SEQ0144	hypothetical phage protein		
			SEQ0145	hypothetical phage protein		
			SEQ0146	hypothetical phage protein		
			SEQ0147	hypothetical phage protein		
			SEQ0148	putative phage RecT family protein		
			SEQ0149	hypothetical phage protein		
			SEQ0150	hypothetical phage protein (pseudogene)		
			SEQ0151	hypothetical phage protein		
			SEQ0152	hypothetical phage protein		
			SEQ0153	hypothetical phage protein		
			SEQ0154	hypothetical phage protein		
			SEQ0155	hypothetical phage protein		
			SEQ0156	hypothetical phage protein		
			SEQ0157	putative phage membrane protein		
			SEQ0158	phage DNA methylase		
			SEQ0159	hypothetical phage protein		
			SEQ0160	hypothetical phage protein		
			SEQ0161	hypothetical phage protein		
			SEQ0162	putative autolysin regulatory protein		
			SEQ0163	phage DNA methylase		
			SEQ0164	hypothetical phage protein		
			SEQ0165	phage terminase		
			SEQ0166	putative minor capsid protein		
			SEQ0167	putative minor capsid protein		
			SEQ0168	hypothetical phage protein		
			SEQ0169	hypothetical phage protein		
			SEQ0170	hypothetical phage protein		
			SEQ0171	putative phage major capsid protein		
			SEQ0172	hypothetical phage protein		
			SEQ0173	hypothetical phage protein		
			SEQ0174	hypothetical phage protein		
			SEQ0175	hypothetical phage protein		
			SEQ0176	hypothetical phage protein		
			SEQ0177	hypothetical phage protein		
			SEQ0178	hypothetical phage protein		
			SEQ0179	putative phage Gp15 protein		

			SEQ0180	putative phage minor tail protein		
			SEQ0181	hypothetical phage protein		
			SEQ0182	hypothetical phage protein		
			SEQ0183	collagen-like repeat phage protein		
			SEQ0184	hypothetical phage protein		
			SEQ0185	hypothetical phage protein		
			SEQ0186	hypothetical phage protein		
			SEQ0187	hypothetical phage protein		
			SEQ0188	putative phage holin		
			SEQ0189	putative phage membrane protein		
			SEQ0190	phage-associated cell wall hydrolase		
			SEQ0191	putative DNA-binding phage protein		
			SEQ0192	hypothetical phage protein		
			SEQ0193	hypothetical phage protein		
			SEQ0194	phage membrane protein		
			SEQ0195	putative DNA-binding phage protein		
			SEQ0196	hypothetical phage protein		
			SEQ0197	hypothetical phage protein		
4			SEQ0214	putative DNA-binding protein (pseudogene)	7	35.62
			SEQ0216*	putative membrane protein		
			SEQ0217*	putative membrane protein		
			SEQ0218*	putative membrane protein		
			SEQ0219*	ABC transporter, ATP-binding protein		
			SEQ0220*	putative membrane protein		
			SEQ0221*	putative membrane protein		
5			SEQ0235	factor H-binding secreted protein Se18.9	5	33.15
			SEQ0236	putative exported protein (pseudogene)		
			SEQ0237	putative membrane protein		
			SEQ0238	putative membrane protein		
			SEQ0239	ABC transporter, ATP-binding protein (pseudogene)		
6			SEQ0281	putative membrane protein	1	30.11
7			SEQ0307	transposase (fragment)	1	35.69
8		similar region in *Szh70*	SEQ0309	conserved hypothetical protein	4	36.73
			SEQ0309a	putative membrane protein		
			SEQ0310	hypothetical protein		
			SEQ0311	hypothetical protein		
9			SEQ0343	hypothetical protein		
10			SEQ0466*	fibronectin-binding protein SFS	1	47.31
11	ICE*Se1* C-terminal region		SEQ0753	putative group II intron reverse transcriptase/maturase	8	34.83
			SEQ0756	putative conjugative transposon regulatory protein		
			SEQ0757	putative modification methylase		
			SEQ0758	putative type II restriction enzyme		
			SEQ0759	putative transposase		
			SEQ0760	putative transposase		
			SEQ0761	conserved hypothetical protein		
			SEQ0762	conserved hypothetical protein		
12	Prophage φSeq2	insertion in *Se4047*	SEQ0787	putative integrase	64	39.24
			SEQ0788	phage membrane protein		
			SEQ0789	putative hypothetical phage protein		
			SEQ0790	putative phage repressor		
			SEQ0791	putative transposase		
			SEQ0792	putative transposase		
			SEQ0793	hypothetical phage protein		
			SEQ0794	putative DNA binding phage protein		
			SEQ0795	hypothetical phage protein		
			SEQ0796	hypothetical phage protein		
			SEQ0797	hypothetical phage protein		
			SEQ0798	hypothetical phage protein		

			SEQ0799	hypothetical phage protein		
			SEQ0800	hypothetical phage protein		
			SEQ0801	putative essential recombination function protein		
			SEQ0802	putative single-strand binding protein		
			SEQ0803	hypothetical phage protein		
			SEQ0804	putative hypothetical phage protein		
			SEQ0805	hypothetical phage protein		
			SEQ0806	phage membrane protein		
			SEQ0807	putative DNA methylase		
			SEQ0808	putative C-5 cytosine-specific DNA methylase (pseudogene)		
			SEQ0810	hypothetical phage protein		
			SEQ0811	phage membrane protein		
			SEQ0812	hypothetical phage protein		
			SEQ0813	hypothetical phage protein		
			SEQ0814	putative autolysin regulatory protein		
			SEQ0815	hypothetical phage protein		
			SEQ0816	hypothetical phage protein		
			SEQ0817	hypothetical phage protein		
			SEQ0818	putative phage HNH endonuclease		
			SEQ0819	putative phage terminase, small subunit		
			SEQ0820	putative phage terminase, large subunit		
			SEQ0821	hypothetical phage protein		
			SEQ0822	putative hypothetical phage protein		
			SEQ0823	putative phage portal protein		
			SEQ0824	putative phage ClpP protease		
			SEQ0825	putative phage major capsid protein		
			SEQ0826	hypothetical phage protein		
			SEQ0827	hypothetical phage protein		
			SEQ0828	hypothetical phage protein		
			SEQ0829	hypothetical phage protein		
			SEQ0830	hypothetical phage protein		
			SEQ0831	major tail protein		
			SEQ0832	hypothetical phage protein		
			SEQ0833	hypothetical phage protein		
			SEQ0834	putative phage tail protein		
			SEQ0835	hypothetical phage protein		
			SEQ0836	hypothetical phage protein		
			SEQ0837	putative phage hyaluronidase		
			SEQ0838	hypothetical phage protein		
			SEQ0839	putative hypothetical phage protein		
			SEQ0840	hypothetical phage protein		
			SEQ0841	hypothetical phage protein		
			SEQ0842	putative phage membrane protein		
			SEQ0843	putative phage membrane protein		
			SEQ0844	putative phage membrane protein		
			SEQ0845	putative phage amidase protein		
			SEQ0846	hypothetical phage protein		
			SEQ0847	hypothetical phage membrane protein		
			SEQ0848	hypothetical phage protein		
			SEQ0849	phospholipase A_2 SlaA		
			SEQ0850	phage membrane protein		
			SEQ0851	hypothetical phage protein		
13			SEQ0902	hypothetical protein		
14		insertion in *Se*4047	SEQ1101	hypothetical protein (fragment)	2	28.1
			SEQ1102	putative conjugative transposon site-specific recombinase		
15			SEQ1183 *diverse	hypothetical protein	1	30.26
16	ICE*Se2*	insertion in *Se*4047	SEQ1229	putative conjugative transposon site-specific recombinase	42	30.6
			SEQ1231	putative DNA-binding protein		

			SEQ1232	hypothetical protein		
			SEQ1233	putative hydrolase EqbN		
			SEQ1234	putative oxidoreductase, EqbM		
			SEQ1235	ABC transporter, ATP-binding membrane protein EqbL		
			SEQ1236	ABC transporter, ATP-binding membrane protein EqbK		
			SEQ1237	ABC transporter, ATP-binding component EqbJ		
			SEQ1238	ABC transporter permease component EqbI		
			SEQ1239	ABC transporter permease component EqbH		
			SEQ1240	equibactin nonribosomal peptide synthase protein EqbG		
			SEQ1241	putative thiazoline reductase EqbF		
			SEQ1242	equibactin nonribosomal peptide synthase protein EqbE		
			SEQ1243	putative salicylate-AMP-ligase EqbD		
			SEQ1244	putative 4'-phosphopantetheinyl transferase EqbC		
			SEQ1245	putative non-ribosomal peptide synthesis thioesterase type II EqbB		
			SEQ1246	iron-dependent repressor EqbA		
			SEQ1247	putative conjugative transposon membrane protein		
			SEQ1249	putative conjugative transposon mobilization protein		
			SEQ1250	putative conjugative transposon mobilization protein		
			SEQ1251	putative conjugative transposon membrane protein		
			SEQ1252	putative conjugative transposon single-strand binding protein		
			SEQ1253	putative conjugative transposon DNA recombination protein		
			SEQ1254	conjugative transposon hypothetical protein		
			SEQ1255	conjugative transposon hypothetical protein		
			SEQ1256	putative DNA topoisomerase		
			SEQ1257	putative abortive infection protein		
			SEQ1258	conjugative transposon hypothetical protein		
			SEQ1259	putative membrane protein		
			SEQ1260	conjugative transposon hypothetical protein		
			SEQ1261	putative conjugative transposon membrane protein		
			SEQ1262	modification DNA methylase		
			SEQ1263	putative conjugal transfer protein		
			SEQ1264	putative conjugative transposon membrane protein		
			SEQ1265	putative conjugative transposon membrane protein		
			SEQ1266	putative conjugative transposon membrane protein		
			SEQ1267	putative conjugative transposon membrane protein		
			SEQ1268	putative conjugative transposon exported protein		
			SEQ1269	putative ABC transporter, ATP-binding protein		
			SEQ1270	putative membrane protein		
			SEQ1271	putative bacteriocin		
			SEQ1272	putative two-component sensor histidine kinase (fragment)		
17		insertion in *Se*4047	SEQ1273	conserved hypothetical protein (fragment)	3	34.47
			SEQ1274	ParB-like nuclease protein		
			SEQ1275	sporulation initiation inhibitor Soj homologue		
18		deletion in	SEQ1328	conserved hypothetical protein	1	28.19

		*Sz*H70 and *Sz*MGCS10565		(pseudogene)		
19		deletion in *Sz*H70	SEQ1517	conserved hypothetical protein	2	37.69
			SEQ1518	hypothetical protein (fragment)		
20			SEQ1654	conserved hypothetical protein	1	30.86
21			SEQ1677a	conserved hypothetical protein (fragment)	1	38.31
22	tRNA		1734575-1734638	tRNA-Gly		
23	prophage φSeq3	insertion in *Se*4047	SEQ1726	hypothetical phage protein	40	39.02
			SEQ1727	exotoxin M precursor SeeM		
			SEQ1728	exotoxin L precursor SeeL		
			SEQ1729	putative phage amidase protein		
			SEQ1730	putative phage membrane protein		
			SEQ1731	putative phage membrane protein		
			SEQ1732	hypothetical phage protein		
			SEQ1733	hypothetical phage protein		
			SEQ1734	hypothetical phage protein		
			SEQ1735	hypothetical phage protein		
			SEQ1736	putative collagen-like repeat protein		
			SEQ1737	hypothetical phage protein		
			SEQ1738	putative phage tail protein		
			SEQ1739	putative phage minor tail protein		
			SEQ1740	hypothetical phage protein		
			SEQ1741	hypothetical phage protein		
			SEQ1742	phage major tail protein		
			SEQ1743	phage major tail protein		
			SEQ1744	hypothetical phage protein		
			SEQ1745	hypothetical phage protein		
			SEQ1746	hypothetical phage protein		
			SEQ1747	hypothetical phage protein		
			SEQ1748	hypothetical phage protein		
			SEQ1749	putative phage capsid protein		
			SEQ1750	hypothetical phage protein		
			SEQ1751	hypothetical phage protein		
			SEQ1752	hypothetical phage protein		
			SEQ1753	hypothetical phage protein		
			SEQ1754	hypothetical phage protein		
			SEQ1755	hypothetical phage protein		
			SEQ1756	hypothetical phage protein		
			SEQ1757	phage Mu protein F like protein		
			SEQ1758	putative phage portal protein		
			SEQ1759	putative phage terminase, large subunit		
			SEQ1760	putative transposase		
			SEQ1761	putative transposase		
			SEQ1762	putative phage repressor protein		
			SEQ1763	hypothetical phage protein		
			SEQ1764	hypothetical phage protein		
			SEQ1765	putative phage integrase		
24			SEQ1853	putative exported protein	1	31.05
25			SEQ1857	putative membrane protein	2	29.72
			SEQ1855a	conserved hypothetical protein		
26	rRNAs / tRNAs		1970594-1976601	16S rRNA, tRNA-Ala, 23S rRNA, 5S rRNA, tRNA-Val, tRNA-Asp, tRNA-Lys, tRNA-Leu, tRNA-Thr, tRNA-Gly, tRNA-Leu, tRNA-Arg, tRNA-Pro		
27	Genomic island		SEQ1971	putative plasmid stabilisation system protein	14	37.18
			SEQ1972	putative plasmid stabilisation system, antitoxin protein		
			SEQ1973	conserved hypothetical protein		
			SEQ1974	type I restriction-modification system R protein		
			SEQ1975	anticodon nuclease		
			SEQ1976	type I restriction-modification system S		

				protein		
			SEQ1977	type I restriction-modification system M protein		
			SEQ1978	conserved hypothetical protein		
			SEQ1979	conserved hypothetical protein		
			SEQ1980	conserved hypothetical protein		
			SEQ1981	conserved hypothetical protein		
			SEQ1982	putative DNA-binding protein		
			SEQ1983	putative DNA-binding protein		
			SEQ1984	putative integrase		
28			SEQ2035*	putative DNA-binding protein (fragment)	1	31.28
29	Prophage φSeq4	insertion in *Se*4047	SEQ2036	exotoxin H precursor SeeH	54	39.18
			SEQ2037	exotoxin I precursor SeeI		
			SEQ2038	putative phage lysin protein		
			SEQ2039	putative phage membrane protein		
			SEQ2040	putative phage holin		
			SEQ2041	phage membrane protein		
			SEQ2042	hypothetical phage protein		
			SEQ2043	hypothetical phage protein		
			SEQ2044	hypothetical phage protein		
			SEQ2045	putative phage hyaluronidase		
			SEQ2046	hypothetical phage protein		
			SEQ2047	hypothetical phage protein		
			SEQ2048	putative phage minor tail protein		
			SEQ2049	hypothetical phage protein		
			SEQ2050	hypothetical phage protein		
			SEQ2051	putative phage major tail protein		
			SEQ2052	hypothetical phage protein		
			SEQ2053	hypothetical phage protein		
			SEQ2054	hypothetical phage protein		
			SEQ2055	hypothetical phage protein		
			SEQ2056	hypothetical phage protein		
			SEQ2057	hypothetical phage protein		
			SEQ2058	hypothetical phage protein		
			SEQ2059	putative phage membrane protein		
			SEQ2060	putative phage minor head protein		
			SEQ2061	putative phage portal protein		
			SEQ2062	putative phage terminase, large subunit		
			SEQ2063	putative phage terminase, small subunit		
			SEQ2064	hypothetical phage protein		
			SEQ2065	putative hypothetical phage protein		
			SEQ2066	hypothetical phage protein		
			SEQ2067	hypothetical phage protein		
			SEQ2068	hypothetical phage protein		
			SEQ2069	hypothetical phage protein		
			SEQ2070	SNF2 family phage protein		
			SEQ2071	hypothetical phage protein		
			SEQ2072	putative phage DNA primase/helicase protein		
			SEQ2073	putative phage DNA polymerase		
			SEQ2074	hypothetical phage protein		
			SEQ2075	hypothetical phage protein		
			SEQ2076	hypothetical phage protein		
			SEQ2077	hypothetical phage protein		
			SEQ2078	hypothetical phage protein (pseudogene)		
			SEQ2079	hypothetical phage protein		
			SEQ2080	hypothetical phage protein		
			SEQ2081	hypothetical phage protein		
			SEQ2082	hypothetical phage protein		
			SEQ2083	putative phage DNA-binding protein		
			SEQ2084	putative phage membrane protein		

			SEQ2085	putative hypothetical phage protein		
			SEQ2086	putative phage DNA-binding protein		
			SEQ2087	hypothetical phage protein		
			SEQ2088	hypothetical phage protein		
			SEQ2089	phage integrase		
30			SEQ2093	hypothetical protein (pseudogene)	3	26.06
			SEQ2094*	major facilitator superfamily protein		
			SEQ2095	putative exported protein		
31		deletion in *Sz*H70 and *Sz*MGCS10565	SEQ2170	ABC transporter membrane protein	4	
			SEQ2171	ABC transporter, ATP-binding subunit		
			SEQ2172	two-component response regulator		
			SEQ2173	putative two-component sensor kinase protein		

* Homologue present in *Sz*MGCS10565

Table A.2B Novel *Sz*H70 DNA loci

No.	Element	Notes	CDS / region	Function	Total no. novel CDS	% GC
1			SZO00830*	putative collagen-like cell surface-anchored protein	1	54.35
2			SZO01230	putative exported protein	1	34.02
3			SZO01450	putative exported protein	1	28.65
4		deletion in *Se*4047	SZO01750*	sorbitol-6-phosphate 2-dehydrogenase SorD	1	43.12
5	genomic island		SZO02230	integrase	8	32.52
			SZO02240	DNA-binding protein		
			SZO02250	conserved hypothetical protein		
			SZO02260	putative membrane protein		
			SZO02270	hypothetical protein		
			SZO02280	conserved hypothetical protein		
			SZO02290	hypothetical protein		
			SZO02300	hypothetical protein		
6			SZO03350 *diverse	putative exported protein	3	31.92
			SZO03360	putative exported protein		
			SZO03370*	putative exported protein		
7	tRNAs		486423 – 486623*	tRNA-Ile, tRNA-Glu		
8		distinct restriction-modification system in *Sz*MGCS10565	SZO04240	putative type I restriction enzyme protein	6	38.39
			SZO04250	hypothetical protein		
			SZO04260	type I restriction modification DNA specificity protein		
			SZO04270	hypothetical protein		
			SZO04280	putative type I restriction enzyme methylase protein		
			SZO04281	hypothetical protein (fragment)		
9			SZO05080	transposase (fragment)	8	30.32
			SZO05081	conserved hypothetical protein		
			SZO05090	putative permease		
			SZO05100	ABC transporter protein		
			SZO05110	putative lantibiotic synthetase protein		
			SZO05120	prolyl oligopeptidase family protein		
			SZO05130	ABC transporter protein		
			SZO05131	conserved hypothetical protein (fragment)		
10		deletion in *Se*4047 mediated by ISSeq3 elements	SZO05580	putative lantibiotic transport ATP-binding protein	12	32.45-43.24
			SZO05590	putative lantibiotic transport protein		
			SZO05600	putative lantibiotic transport protein		
			SZO05610*	ABC transporter protein		
			SZO05620*	putative membrane protein		
			SZO05630*	putative membrane protein		
			SZO05640	dhaKLM operon coactivator DhaQ		
			SZO05650	dhaKLM operon transcriptional activator DhaS		
			SZO05660	PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit DhaK		
			SZO05670	PTS-dependent dihydroxyacetone kinase, ADP-binding subunit DhaL		
			SZO05680	sugar phosphotransferase system (PTS), IIC component		
			SZO05690	glycerol uptake facilitator protein GlpF		
11		similar region in *Se*4047 and *Sz*MGCS10565	SZO05820 *diverse	putative exported protein	5	36.37
			SZO05830 *diverse	putative exported protein		
			SZO05840	putative exported protein (pseudogene)		
			SZO05850	putative exported protein		

			SZO05860*diverse	conserved hypothetical protein		
12		deletion in *Se*4047 and *Sz*MGCS10565	SZO06070	phosphate transport system protein PhoU (pseudogene)	1	40.51
13		deletion in *Se*4047 mediated by ISSeq3 elements	SZO06280	hypothetical protein	6	43.47
			SZO06290 *diverse	putative membrane protein		
			SZO06300 *diverse	conserved hypothetical protein		
			SZO06310 *diverse	conserved hypothetical protein		
			SZO06320 *diverse	putative membrane protein		
			SZO06330	conserved hypothetical protein		
14	genomic island		SZO07580	replication initiator protein (fragment)	12	31.26
			SZO07581	conserved hypothetical protein (fragment)		
			SZO07590	mobilisation protein		
			SZO07600	putative relaxase (fragment)		
			SZO07620	sensor histidine kinase		
			SZO07630	response regulator protein		
			SZO07640	putative membrane protein		
			SZO07660	lantibiotic biosynthesis protein		
			SZO07670	lantibiotic transport/processing ATP-binding protein		
			SZO07680	putative lantibiotic ABC transporter protein		
			SZO07690	putative lantibiotic ABC transporter protein		
			SZO07700	putative lantibiotic ABC transporter protein		
15		similar region in *Se*4047 and *Sz*MGCS10565	SZO08010*	putative membrane protein	2	34.11
			SZO08020*	putative membrane protein		
			SZO08030*	putative membrane protein		
			SZO08040*	putative membrane protein		
			SZO08050	putative membrane protein		
16		deletion in *Se*4047	SZO08550*	putative recombinase	2	37.64
			SZO08560*	putative cell surface-anchored protein		
17	genomic island		SZO09570	conserved hypothetical protein	9	30.25
			SZO09580	putative protein kinase		
			SZO09590	hypothetical protein		
			SZO09600	conserved hypothetical protein		
			SZO09610	conserved hypothetical protein		
			SZO09620	hypothetical protein		
			SZO09630	type II restriction enzyme-methylase		
			SZO09640	type II restriction enzyme and methylase		
			SZO09650	hypothetical protein		
18		deletion in *Se*4047 mediated by ISSeq3 elements	SZO09940*	bacteriocin-associated membrane protein	6	35.92
			SZO09941*	putative bacteriocin		
			SZO09950*	putative thioredoxin		
			SZO09951*	putative bacteriocin		
			SZO09952*	putative bacteriocin		
			SZO09960*	putative membrane protein		
19			SZO10340*	conserved hypothetical protein	11	32.24
			SZO10361*	conserved hypothetical protein (fragment)		
			SZO10360*	putative exported protein		
			SZO10370*	conserved hypothetical protein		
			SZO10380*	conserved hypothetical protein		
			SZO10390	conserved hypothetical protein		
			SZO10400	putative exported protein		
			SZO10410	conserved hypothetical protein (fragment)		
			SZO10420	putative exported protein		
			SZO10430*	conserved hypothetical protein		

				(pseudogene)		
			SZO10440*	putative membrane protein		
20		deletion in *Se*4047 mediated by recombination between ISSeq3 and ISSeq5 elements	SZO12330*	putative ABC transporter, ATP-binding/permease protein	8	33.77
			SZO12340	putative membrane protein		
			SZO12350	ABC-2 type transporter protein		
			SZO12360	ABC transporter protein		
			SZO12370	hypothetical protein		
			SZO12380	radical SAM superfamily protein		
			SZO12390	radical SAM superfamily protein		
			SZO12400	putative DNA-binding protein		
21	ICE*Sz1* C-terminus		SZO12560	conserved hypothetical protein	17	34.57
			SZO12570	conserved hypothetical protein (fragment)		
			SZO12580	conserved hypothetical protein (fragment)		
			SZO12590	conserved hypothetical protein (fragment)		
			SZO12600	putative DNA-binding protein		
			SZO12660	DNA topoisomerase (fragment)		
			SZO12670	putative replication initiation protein (fragment)		
			SZO12680	putative membrane protein		
			SZO12690	conserved hypothetical protein		
			SZO12700	site-specific recombinase		
			SZO12710	conserved hypothetical protein		
			SZO12720	radical SAM superfamily protein		
			SZO12730	conserved hypothetical protein		
			SZO12740	conserved hypothetical protein		
			SZO12750	putative DNA-binding protein		
			SZO12760	conserved hypothetical protein		
			SZO12770	relaxase/mobilisation protein		
22	ICE*Sz1* N-terminal region		SZO12900	group II intron reverse transcriptase maturase (pseudogene)	2	34.23
			SZO12930	ABC transporter protein		
23			SZO13420	putative exported protein	1	34.68
24			SZO14080*	response regulatory protein	4	38.04
			SZO14090*	sensor histidine kinase		
			SZO14100*	ABC-2 type transporter protein		
			SZO14110*	ABC transporter protein		
25	CRISPR locus	deletion in *Se*4047 mediated by ISSeq11 elements	1604884 – 1606107 *distinct repeats	CRISPR repeats	24	40.56
			SZO14370*	CRISPR-associated protein Cas2		
			SZO14380*	CRISPR-associated protein Cas1		
			SZO14390*	CRISPR-associated protein Cas4		
			SZO14400*	CRISPR-associated protein		
			SZO14410*	CRISPR-associated protein		
			SZO14420*	CRISPR-associated protein		
			SZO14430*	CRISPR-associated helicase		
			SZO14440*	hypothetical protein		
			SZO14450*	hypothetical protein		
			SZO14460	hypothetical protein		
			SZO14471*	conserved hypothetical protein (fragment)		
			SZO14470*	conserved hypothetical protein		
			SZO14480	hypothetical protein		
			SZO14490*	conserved hypothetical protein		
			SZO14500*	hypothetical protein		
			SZO14510*	putative membrane protein		
			SZO14520	conserved hypothetical protein		
			SZO14530	conserved hypothetical protein		
			SZO14540	hypothetical protein		
			SZO14550*	hypothetical protein		
			SZO14560	putative membrane protein		

			SZO14570*	hypothetical protein		
			SZO14580*	hypothetical protein		
			SZO14590*	putative membrane protein		
26	ESAT-6-like locus	deletion in *Se*4047 mediated by ISSeq3 elements	SZO14660*	ESAT-6 secretion system protein EssA	4	42.29
			SZO14670*	ESAT-6 secretion system protein EsaA		
			SZO14680*	ESAT-6 secreted protein EsxA		
			SZO14690*	branched-chain amino acid transport system carrier protein		
27			SZO14721*	conserved hypothetical protein (fragment)	8	33.1
			SZO14730*	transport protein ComB		
			SZO14740*	transport/processing ATP-binding protein ComA		
			SZO14744*	putative competence stimulating peptide		
			SZO14743*	hypothetical protein		
			SZO14742*	putative competence stimulating peptide		
			SZO14741*	conserved hypothetical protein		
			SZO14750*	DNA-binding protein		
28		deletion in *Se*4047 and *Sz*MGCS10565	SZO14870	putative exported protein	1	36.51
29		deletion in *Se*4047	SZO15030*	ABC transporter, ATP-binding protein	4	41.1
			SZO15040*	putative membrane protein		
			SZO15050*	putative membrane protein		
			SZO15060*	conserved hypothetical protein		
30		deletion in *Se*4047	SZO15110*	D-ribose-binding protein precursor RbsD	6	43.95
			SZO15120*	ribose transport system permease protein RbsC		
			SZO15130*	ribose import ATP-binding protein RbsA		
			SZO15140*	high affinity ribose transport protein RbsD		
			SZO15150*	ribokinase RbsK		
			SZO15160*	ribose operon repressor RbsR		
31		deletion in *Se*4047	SZO15220*	6-phospho-beta-galactosidase LacG	5	44.11
			SZO15230*	sugar phosphotransferase system (PTS), lactose-specific IICB component LacE		
			SZO15240*	sugar phosphotransferase system (PTS), lactose-specific phosphotransferase enzyme IIA component LacF		
			SZO15250*	transcription antiterminator LacT		
			SZO15260*	tagatose 1,6-diphosphate aldolase LacD		
32		similar region in *Se*4047	SZO17200	hypothetical protein	12	37.97
			SZO17210	hypothetical protein		
			SZO17220	hypothetical protein		
			SZO17230	hypothetical protein		
			SZO17240	hypothetical protein		
			SZO17250	hypothetical protein		
			SZO17260	conserved hypothetical protein		
			SZO17270	hypothetical protein		
			SZO17280	hypothetical protein		
			SZO17290	conserved hypothetical protein (fragment)		
			SZO17300*	conserved hypothetical protein		
			SZO17310	conserved hypothetical protein		
33			SZO17550	conserved hypothetical protein	2	29.63
			SZO17560	hypothetical protein		
34	ICE*Sz2*	insertion in *Sz*H70	SZO17561	relaxase (fragment)	41	38.64
			SZO17570	group II intron-encoded protein (pseudogene)		

			SZO17580	putative relaxase (fragment)		
			SZO17590	mobilisation protein		
			SZO17600	conserved hypothetical protein		
			SZO17610	DNA-binding protein		
			SZO17620	conserved hypothetical protein		
			SZO17630	conserved hypothetical protein		
			SZO17640	conserved hypothetical protein		
			SZO17650	conserved hypothetical protein		
			SZO17660	conserved hypothetical protein (pseudogene)		
			SZO17680	putative exported protein		
			SZO17690	multi antimicrobial extrusion (MATE) family transporter		
			SZO17700	conserved hypothetical protein		
			SZO17710	radical SAM superfamily protein		
			SZO17730	transcriptional regulator		
			SZO17740	putative membrane protein		
			SZO17750	conserved hypothetical protein		
			SZO17760	putative helicase		
			SZO17770	putative group II intron reverse transcriptase/maturase		
			SZO17790	conserved hypothetical protein		
			SZO17800	putative glucan-binding protein		
			SZO17810	hypothetical protein		
			SZO17820	conserved hypothetical protein		
			SZO17830	conserved hypothetical protein		
			SZO17840	putative amidase (fragment)		
			SZO17850	putative peptidase (fragment)		
			SZO17860	conserved hypothetical protein (pseudogene)		
			SZO17870	putative group II intron reverse transcriptase/maturase		
			SZO17890	conserved hypothetical protein (pseudogene)		
			SZO17900	putative membrane protein		
			SZO17910	putative membrane protein		
			SZO17920	TraG family protein (pseudogene)		
			SZO17940	conserved hypothetical protein		
			SZO17950	abortive infection protein		
			SZO17960	conserved hypothetical protein		
			SZO17970	conserved hypothetical protein (fragment)		
			SZO17980	conserved hypothetical protein (fragment)		
			SZO17990	C-5 cytosine-specific DNA methylase		
			SZO18000	replication initiator A protein		
			SZO18010	conserved hypothetical protein		
35		deletion in *Se*4047 mediated by ISSeq3, 2 distinct pilus clusters in *Sz*MGCS10565	SZO18270	sortase SrtC.2	8	37.91
			SZO18280	sortase SrtC.3		
			SZO18290	sortase SrtC.4		
			SZO18300	putative exported protein		
			SZO18310	putative ancillary pilus subunit		
			SZO18320	putative backbone pilus subunit		
			SZO18330	putative pilus subunit		
			SZO18340*	AraC family regulatory protein		
36	genomic island	insertion in *Sz*H70 and *Sz*MGCS10565 into the end of *rpsD*	SZO19050*	plasmid stabilization system protein	4	36.88
			SZO19060*	conserved hypothetical protein		
			SZO19061*	phage hypothetical protein (pseudogene)		
			SZO19062*	phage hypothetical protein (fragment)		
37		similar region in *Se*4047	SZO19300*	putative membrane protein	1	34.61
			SZO19310*	putative membrane protein		
			SZO19320*	putative membrane protein		
			SZO19330	putative membrane protein		

* Homologue present in *Sz*MGCS10565

Table A.2C Diversified regions of the *Sz*H70 and *Se*4047 genomes

No.	*Se*4047	*Sz*H70	% Similarity §	Function
1	SEQ0084	SZO00790	78.9	hypothetical protein
2	SEQ0090	SZO00840	66.7	putative collagen-like cell surface-anchored protein SclG
3	SEQ0121	SZO01080	77.0	putative membrane protein
4	SEQ0202	SZO01211†	77.8	putative exported protein
	SEQ0203	SZO01220	88.8	putative exported protein
5	SEQ0256	SZO01590	81.2	putative cell surface-anchored protein
	SEQ0260	SZO01630	84.7	putative collagen-like cell surface-anchored protein SclH
6	SEQ0280	SZO17540	76.5	putative collagen-like cell surface-anchored protein SclD
7	SEQ0308	SZO17320	71.6	membrane protein
8	SEQ0367	SZO16670	76.2	MerR family regulatory protein
9	SEQ0375 †	SZO16630	71.5	collagen- and fibronectin-binding protein Fne/Fnz
10	SEQ0402	SZO16370	71.9	putative cell surface-anchored protein
11	SEQ0432	SZO16140	90.7	endoglycosidase EndoS
12	SEQ0521	SZO15350	91.7	putative negative regulator of copper transport operon CopY
13	SEQ0527	SZO15290	93.6	galactose-6-phosphate isomerase LacA subunit
14	SEQ0555	SZO14890	45.1	putative collagen and fibronectin-binding cell surface-anchored protein FneE
15	SEQ0566 †	SZO14790	59.3	Szp/SzPSe-like cell surface-anchored protein
16	SEQ0633	SZO13850	80.1	putative collagen-like surface-anchored protein SclE
17	SEQ0855	SZO12230	83.3	putative collagen-like surface-anchored protein SclF
18	SEQ0896	SZO11861	63.4	putative exported protein
19	SEQ0933	SZO11530	84.2	fibrinogen-binding cell surface-anchored protein SzPSe; adhesin and hypervariable protective antigen Szp
20	SEQ0938	SZO11480	81.3	Mac family IgG endopeptidase antiphagocytic and neutrophil-binding protein IdeE/IdeZ
	SEQ0939	SZO11470	75.5	putative cell surface-anchored protein
21	SEQ1051	SZO10440	51.0	putative membrane protein
22	SEQ1510	SZO06420	82.2	conserved hypothetical protein
	SEQ1511	SZO06410†	88.5	hypothetical protein
23	SEQ1543	SZO06080	85.2	putative exported protein
24	SEQ1564	SZO05860	58.7	conserved hypothetical protein
	SEQ1565	SZO05820	73.5	putative membrane protein
25	SEQ1606	SZO05380	67.3	putative collagen-binding collagen-like surface-anchored protein FneC
26	SEQ1607 †	SZO05350	48.3	putative collagen-binding surface-anchored protein FneD
27	SEQ1649	SZO04910	63.1	putative collagen-like surface-anchored protein FneF
28	SEQ1670	SZO04720	44.9	putative exported protein
29	SEQ1817	SZO03720	79.3	collagen-like cell surface-anchored protein SclI
30	SEQ1941	SZO02570	74.6	putative membrane protein
31	SEQ1999	SZO02080	85.0	fibronectin and collagen-binding cell surface anchored protein FneB/Fnz2
32	SEQ2017	SZO01900	74.7	antiphagocytic cell surface-anchored fibrinogen- and IgG Fc-binding protein
33	SEQ2101	SZO18110	69.1	putative collagen-like surface-anchored protein SclC

§ Needleman-Wunsch global alignment of amino acid sequences. † Pseudogene or partial gene. Regions 1 to 33 are also diverse in *Sz*MGCS10565 relative to *Se*4047 and/or *Sz*H70 (a homologous CDS is absent in *Sz*MGCS10565 in regions 7, 23 and 30).

Table A.2D *Se*4047 pseudogenes (includes partial genes)

No.	*Se*4047	*Sz*H70 ortholog	Intact in *Sz*H70	Function
1	SEQ0036†	SZO00360	no	conserved hypothetical protein
2	SEQ0036a †	SZO00370	no	conserved hypothetical protein
3	SEQ0038	SZO00390	yes	putative DNA-binding protein
4	SEQ0048	not present	-	ThiF family protein
5	SEQ0083	SZO00780	no	transposase
6	SEQ0085	SZO00800	yes	putative membrane protein
7	SEQ0100†	not present	-	putative transposase
8	SEQ0115	not present	-	transposase
9	SEQ0150 (φSeq1)	not present	-	hypothetical phage protein
10	SEQ0214	not present	-	putative DNA-binding protein
11	SEQ0232†	SZO01430	yes	putative cell wall anchored protein Spa
12	SEQ0236	not present	-	putative exported protein
13	SEQ0239	not present	-	ABC transporter, ATP-binding protein
14	SEQ0273†	SZO17520	yes	putative rhomboid family membrane protein
15	SEQ0275	not present	-	transposase
16	SEQ0275	not present	-	transposase
17	SEQ0307†	not present	-	transposase
18	SEQ0309a †	not present	-	putative membrane protein
19	SEQ0313	SZO17180	yes	hypothetical protein
20	SEQ0320	SZO17120	yes	putative N-acetylmannosamine-6-phosphate 2-epimerase
21	SEQ0328†	SZO17040	yes	putative acetyl xylan esterase (AXE1) family protein
22	SEQ0370	SZO16650	yes	response regulator SaeR
23	SEQ0375	SZ016630	yes	Fne collagen and fibronectin-binding protein
24	SEQ0380	SZO16601	yes	putative DNA-binding protein
25	SEQ0405†	SZO16340	no	putative hypothetical phage protein
26	SEQ0424	SZO16200	yes	conserved hypothetical protein
27	SEQ0426	SZO16190	yes	conserved hypothetical protein
28	SEQ0438	not present	-	transposase
29	SEQ0443	SZO16070	yes	Shr heme binding cell surface protein
30	SEQ0495	SZO15570	yes	putative haloacid dehalogenase-like hydrolase
31	SEQ0509a †	SZO15450	no	putative integrase
32	SEQ0509b †	SZO15440	no	putative transposase
33	SEQ0511†	SZO15430	no	putative membrane protein
34	SEQ0514†	SZO15420	no	conserved hypothetical protein
35	SEQ0530†	SZO15260	yes	putative tagatose 1,6-diphosphate aldolase
36	SEQ0531†	SZO15220	yes	6-phospho-beta-galactosidase LacG
37	SEQ0541†	SZO15030	yes	ABC transporter, ATP-binding protein
38	SEQ0556†	not present	-	Scl surface protein
39	SEQ0566	SZO14790	yes	M-like surface-anchored protein
40	SEQ0573†	not present	-	transposase
41	SEQ0576†	SZO14660	yes	ESAT-6 secretion system protein EssA
42	SEQ0578	SZO14651	yes	ESAT-6 secretion system protein EsaB
43	SEQ0583a †	SZO14590	yes	putative membrane protein
44	SEQ0584†	SZO14360	yes	transposase
45	SEQ0585†	SZO14350	yes	transposase
46	SEQ0663	SZO13550	yes	transposase
47	SEQ0685	SZO13370	yes	putative competence protein CoiA-like family protein
48	SEQ0738	SZO12890	yes	putative membrane protein on ICESe1
49	SEQ0733	SZO12940	no	putative conjugal transfer protein
50	SEQ0746†	SZO12820	yes	putative DNA topoisomerase on ICESe1
51	SEQ0749†	SZO12660	yes	putative DNA topoisomerase on ICESe1
52	SEQ0763†	SZO12770	yes	putative relaxase/mobilization protein end of ICESe1 but not end of ICEsz1
53	SEQ0774	SZO12430	yes	putative oligopeptidase

54	SEQ0780†	SZO12330	yes	putative ABC transporter, ATP-binding/permease protein
55	SEQ0808 (φSeq2)	not present	-	putative C-5 cytosine-specific DNA methylase
56	SEQ0866	SZO12130	yes	putative beta-galactosidase precursor
57	SEQ0871	SZO12110	yes	sugar phosphotransferase system (PTS), sorbose-specific family, IIC component
58	SEQ0899†	SZO11830	no	transposase
59	SEQ0909	SZO11740	yes	putative 1,4-alpha-glucan branching enzyme
60	SEQ0934	SZO11520	yes	TetR family regulatory protein
61	SEQ0943†	SZO11420	no	transposase
62	SEQ0986	not present	-	transposase
63	SEQ1007	not present	-	transposase
64	SEQ1064	SZO10210	yes	putative ABC transporter permease protein
65	SEQ1068	not present	-	transposase
66	SEQ1077	SZO10150	yes	putative cell surface-anchored C5A peptidase precursor
67	SEQ1088	SZO10050	yes	conserved hypothetical protein
68	SEQ1098†	SZO09960	yes	putative membrane protein
69	SEQ1101†	not present	-	hypothetical protein
70	SEQ1103†	SZO08820	yes	putative ABC transporter ATP-binding membrane protein
71	SEQ1149	SZO09270	yes	putative ABC-type glycine betaine transport system ATP-binding protein
72	SEQ1151	SZO09280	yes	putative glycine betaine ABC transporter permease and substrate binding protein
73	SEQ1155	SZO09310	yes	ApbE family protein
74	SEQ1189	SZO09690	yes	putative ABC transporter permease
75	SEQ1209	SZO09840	yes	putative ammonia monooxygenase
76	SEQ1223	SZO09930	yes	putative ABC transporter ATP-binding protein
77	SEQ1225†	SZO09940	yes	bacteriocin-associated membrane protein
78	SEQ1226†	SZO09950	yes	putative thioredoxin (bacteriocin cluster)
79	SEQ1272†	not present	-	putative two-component sensor histidine kinase (on start of ICESe2)
80	SEQ1273†	not present	-	hypothetical protein (upstream of ICESe2)
81	SEQ1276†	SZO08820	yes	putative ABC transporter ATP-binding membrane protein
82	SEQ1281	SZO08770	yes	putative ABC transporter, ATP-binding/permease protein
83	SEQ1284	SZO08750	yes	putative ABC transporter ATP-binding protein
84	SEQ1307a †	SZO08560	yes	putative cell surface-anchored protein
85	SEQ1328	not present	-	conserved hypothetical protein
86	SEQ1364†	SZO07970	no	putative uncharacterized protein
87	SEQ1380†	SZO07810	no	putative membrane protein (bacteriocin associated)
88	SEQ1383†	SZO07780	no	putative bacteriocin transport/processing ATP-binding protein
89	SEQ1385	SZO07770	no	putative bacteriocin secretion protein
90	SEQ1387	SZO07750	no	putative transposase
91	SEQ1467	SZO06800	yes	aminopeptidase PepS
92	SEQ1479	SZO06680	yes	hyaluronate lyase precursor HysA
93	SEQ1492	SZO06560	yes	conserved hypothetical protein
94	SEQ1494	SZO06550	yes	conserved hypothetical protein
95	SEQ1507	SZO06440	yes	putative membrane (in operon with RNA polymerase sigma factor protein)
96	SEQ1518†	SZO06361	no	hypothetical protein
97	SEQ1523	SZO06260	yes	major facilitator superfamily protein
98	SEQ1528†	SZO06220	no	transposase
99	SEQ1573	not present	-	putative transposase
100	SEQ1585†	SZO05580	yes	putative lantibiotic transport ATP-binding protein
101	SEQ1588a †	SZO05541	no	conserved hypothetical protein
102	SEQ1589†	SZO05540	no	putative transposase
103	SEQ1590†	SZO05530	no	putative transposase
104	SEQ1607	SZO05350	yes	collagen-binding surface-anchored protein
105	SEQ1613	SZO05300	yes	sugar phosphotransferase system (PTS), IIC component
106	SEQ1632†	SZO05140	no	conserved hypothetical protein
107	SEQ1667	SZO04750	yes	conserved hypothetical protein
108	SEQ1671	SZO04710	yes	putative exported protein
109	SEQ1677†	SZO04650	no	conserved hypothetical protein
110	SEQ1677a	not present	-	conserved hypothetical protein

	†			
111	SEQ1699	SZO04441	no	putative exported protein
112	SEQ1720†	SZO04200	no	transposase
113	SEQ1724	SZO04180	yes	putative late competence protein ComFC
114	SEQ1725	SZO04170	yes	putative late competence protein ComFA
115	SEQ1810	SZO03760	yes	putative exported protein
116	SEQ1824a †	SZO03651	no	transposase
117	SEQ1854†	SZO03390	yes	conserved hypothetical protein
118	SEQ1855a †	not present	-	conserved hypothetical protein
119	SEQ1873	not present	-	transposase
120	SEQ1924	not present	-	transposase
121	SEQ1942†	SZO02560	no	integrase
122	SEQ1966†	SZO02330	no	transposase
123	SEQ1968†	SZO02310	no	transposase
124	SEQ2010	SZO01970	yes	conserved hypothetical protein
125	SEQ2035†	not present	-	putative DNA-binding protein
126	SEQ2078 (φSeq4)	not present	-	hypothetical phage protein
127	SEQ2093	not present	-	hypothetical protein
128	SEQ2099	SZO18092	yes	conserved hypothetical protein
129	SEQ2117	not present	-	transposase
130	SEQ2119†	SZO18340	yes	AraC family regulatory protein
131	SEQ2122	SZO18370	yes	deoxynucleoside kinase
132	SEQ2138	SZO18520	yes	putative membrane protein
133	SEQ2184	SZO18940	yes	conserved hypothetical protein
134	SEQ2192	SZO18990	yes	putative MutT/NUDIX hydrolase family protein
135	SEQ2225	not present	-	transposase
136	SEQ2227	SZO19320	yes	putative membrane protein

† indicates partial gene

Table A.2E *Sz*H70 pseudogenes (includes partial genes)

No.	*Sz*H70	*Se*4047 ortholog	Intact in *Se*4047	Function
1	SZO00360†	SEQ0036	no	conserved hypothetical protein
2	SZO00370†	SEQ0036a	no	conserved hypothetical protein
3	SZO00780†	SEQ0083	no	transposase
4	SZO01211	SEQ0202	yes	putative exported protein
5	SZO01310	SEQ0213	yes	putative mutator protein MutX
6	SZO01811†	SEQ2027	yes	NUDIX hydrolase
7	SZO01890	SEQ2018	yes	AraC family regulatory protein
8	SZO02310†	SEQ1968	no	transposase
9	SZO02330†	SEQ1966	no	transposase
10	SZO02560†	SEQ1942	no	integrase
11	SZO03651†	SEQ1824a	no	transposase
12	SZO04200†	SEQ1720	no	transposase
13	SZO04281†	not present	-	hypothetical protein
14	SZO04441	SEQ1699	no	putative exported protein
15	SZO04650†	SEQ1677	no	conserved hypothetical protein
16	SZO05080†	not present	-	transposase
17	SZO05131†	not present	-	conserved hypothetical protein
18	SZO05140†	SEQ1632	no	conserved hypothetical protein
19	SZO05460	SEQ1598	yes	putative ABC transporter, ATP-binding/permease protein
20	SZO05530†	SEQ1590	no	putative transposase
21	SZO05540†	SEQ1589	no	putative transposase
22	SZO05541†	SEQ1588a	no	transposase
23	SZO05550	SEQ1588	yes	putative lantibiotic leader peptide processing serine protease
24	SZO05840	not present	-	putative exported protein
25	SZO06070	not present	-	phosphate transport system protein PhoU
26	SZO06220†	SEQ1528	no	transposase
27	SZO06320†	not present	-	putative membrane protein
28	SZO06360†	SEQ1516	yes	hypothetical protein
29	SZO06361†	SEQ1518	no	hypothetical protein
30	SZO06380	SEQ1514	yes	hypothetical protein
31	SZO06410†	SEQ1511	yes	hypothetical protein
32	SZO07580†	not present	-	replication initiator protein (on genomic island)
33	SZO07581†	not present	-	conserved hypothetical protein
34	SZO07600†	not present	-	putative relaxase
35	SZO07660	not present	-	lantibiotic biosynthesis protein
36	SZO07710†	SEQ1389	yes	transposase
37	SZO07750	SEQ1387	no	putative transposase
38	SZO07770	SEQ1385	no	putative bacteriocin secretion protein
39	SZO07780†	SEQ1383	no	putative bacteriocin transport/processing ATP-binding protein
40	SZO07790	SEQ1382	yes	putative bacteriocin
41	SZO07810†	SEQ1380	no	putative membrane protein (bacteriocin associated)
42	SZO07970†	SEQ1364	no	putative uncharacterized protein
43	SZO08000	SEQ1361	yes	putative membrane protein
44	SZO10201†	SEQ1066	yes	transposase
45	SZO10361†	not present	-	conserved hypothetical protein
46	SZO10410†	not present	-	conserved hypothetical protein
47	SZO10430	not present	-	conserved hypothetical protein
48	SZO11420†	SEQ0943	no	transposase
49	SZO11830†	SEQ0899	no	transposase
50	SZO12460	SEQ0771	yes	putative membrane protein
51	SZO12570†	not present	-	conserved hypothetical protein
52	SZO12580†	not present	-	conserved hypothetical protein
53	SZO12590†	not present	-	conserved hypothetical protein
54	SZO12620†	SEQ0752	no	putative conjugative transposon DNA recombination protein
55	SZO12630†	SEQ0752	no	putative conjugative transposon DNA recombination protein
56	SZO12640	SEQ0751	yes	DNA-binding protein

57	SZO12660†	not present	-	DNA topoisomerase
58	SZO12670†	not present	-	putative replication initiation protein
59	SZO12900	not present	-	group II intron reverse transcriptase maturase
60	SZO12921	SEQ0737	yes	Putative conjugative transposon membrane protein
61	SZO12940	SEQ0733	no	putative conjugal transfer protein
62	SZO14260	SEQ0595	yes	arginine repressor
63	SZO14360	SEQ0584	no	putative transposase
64	SZO14471†	not present	-	conserved hypothetical protein
65	SZO14721†	not present	-	conserved hypothetical protein
66	SZO15430†	SEQ0511	no	putative membrane protein
67	SZO15420†	SEQ0514	no	conserved hypothetical protein
68	SZO15440†	SEQ0509b	no	putative transposase
69	SZO15450†	SEQ0509a	no	putative integrase
70	SZO16340†	SEQ0405	no	putative hypothetical phage protein
71	SZO16870	SEQ0346	yes	acid phosphatase precursor LppC
72	SZO17290†	SEQ0308	yes	conserved hypothetical protein
73	SZO17330†	not present	-	transposase
74	SZO17561†	not present	-	relaxase
75	SZO17570	not present	-	group II intron-encoded protein
76	SZO17580†	not present	-	putative relaxase
77	SZO17660	not present	-	conserved hypothetical protein
78	SZO17840†	not present	-	putative amidase
79	SZO17850†	not present	-	putative peptidase
80	SZO17860	not present	-	conserved hypothetical protein
81	SZO17890	not present	-	conserved hypothetical protein
82	SZO17920	not present	-	TraG family protein
83	SZO17970†	not present	-	conserved hypothetical protein
84	SZO17980†	not present	-	conserved hypothetical protein
85	SZO18020†	SEQ1389	yes	putative transposase
86	SZO18030†	SEQ1389	yes	putative transposase
87	SZO18031† and SZO01811†	SEQ2027	yes	NUDIX hydrolase
88	SZO18820†	SEQ2170	yes	ABC transporter membrane protein
89	SZO18970†	SEQ2190	yes	putative cell surface-anchored protein
90	SZO19061	not present	-	phage hypothetical protein
91	SZO19062†	not present	-	phage hypothetical protein

† indicates partial gene

Table A.3 Composition of chemically defined medium

Component	Concentration (g/L)
Adenine Sulfate, $2H_2O$	0.020000
DL-Alanine	0.100000
L-Arginine, FB	0.100000
L-Asparagine, Anhydrous	0.100000
L-Aspartic Acid	0.100000
Biotin	0.000200
Calcium Chloride, Anhydrous *	0.005066
D-Calcium Pantothenate	0.002000
Cyanocobalamin	0.000100
L-cystine, 2HCl	0.065200
Dextrose, Anhydrous	10.000000
Ferric Citrate *	0.003040
Ferrous Sulfate, $7H_2O$ *	0.005000
Folic Acid	0.000800
L-Glutamic Acid	0.100000
L-Glutamine	0.200000
Glycine	0.100000
Guanine, HCl, H_2O	0.020000
L-Histidine, FB	0.100000
Hydroxy L-Proline	0.100000
L-Isoleucine	0.100000
L-Leucine	0.100000
L-Lysine, HCl	0.124900
Magnesium Sulfate, heptahydrate *	0.700000
Manganous Sulfate, H_2O *	0.005850
L-Methionine	0.100000
Niacinamide	0.001000
Beta-NAD, 3 H_2O	0.002500
Para-Aminobenzoic Acid	0.000200
L-Phenylalanine	0.100000
Potassium Phosphate, Dibasic, Anhydrous	0.200000
Potassium Phosphate, Monobasic, Anhydrous	1.000000
L-Proline	0.100000
Pyridoxal, HCl	0.001000
Pyridoxamine, 2HCl	0.001000
Riboflavin	0.002000
L-Serine	0.100000
Sodium Acetate, Anhydrous	2.712600
Sodium Phosphate, Dibasic, Anhydrous	7.350000
Sodium Phosphate, Monobasic, H_2O, ACS	3.195000
Thiamine, HCl	0.001000
L-Threonine	0.200000
L-Tryptophan	0.100000
L-Tyrosine, 2Na, $2H_2O$	0.144200

Table A.3 continued

Component	Concentration (g/L)
Uracil	0.020000
L-Valine	0.100000

Composition of chemically defined medium (CDM) purchased in powder form from SAFC Biosciences. * Reagents from Sigma added separately.

Table A.4 Identity of *S. equi* and *S. zooepidemicus* isolates studied by Southern blot and PCR

Lane	Species	PCR type or SeM allele	ST	Strain ID	Disease/Source	Animal	Year
1	*Se*	D1, SeM-3	179	4047	Strangles	Horse	1990
2	Se	SeM-9	179	7344	Strangles	Horse	2003
3	Se	SeM-6	NA	7350	Strangles	Horse	2003
4	Se	SeM-6	179	1350	Strangles	Horse	2003
5	Se	SeM-11	179	7364	Strangles	Horse	2003
6	Se	SeM-2	179	CF32	Strangles*	Horse	1981
7	Se	SeM-5	179	B23 7325	Strangles	Horse	2003
8	Se	SeM-1	NA	6073	Strangles	Horse	1998
9	Se	NA	179	8229	Strangles	Dog	2004
10	Sz	D1/1	23	H25	Lower respiratory	Horse	2000
11	Sz	D1/1	23	H61	Upper respiratory	Horse	2000
12	Sz	D1/1	NA	D53	Lower respiratory	Horse	2000
13	Sz	D1/1	8	H8	Lower respiratory	Horse	2000
14	Sz	D1/3	NA	D24	Lower respiratory	Horse	2000
15	Sz	D1/4	51	B23 6074	Uterine infection/abortion	Horse	2003
16	Sz	D1/5	33	H50	Upper respiratory	Horse	2000
17	Sz	D1/5	27	H15	Lower respiratory	Horse	2000
18	Sz	D1/u	NA	D1a	Lower respiratory	Horse	2000
19	Sz	D1/u	NA	D84	Upper respiratory	Horse	2000
20	Sz	D1/u	49	B23 6378	Uterine infection/abortion	Horse	2003
21	Sz	A1/1	1	7271	Uterine infection/abortion	Horse	2003
22	Sz	A1/2	NA	P22	Lower respiratory	Horse	2000
23	Sz	A1/2	NA	BHS 32	Lower respiratory	Dog	2001
24	Sz	A1/2	22	BHS 53	Lower respiratory	Dog	2001
25	Sz	A1/3	30	H9	Lower respiratory	Horse	2000
26	Sz	A1/4	NA	D36	Lower respiratory	Horse	2000
27	Sz	A1/4	13	B26 O863	Uterine infection/abortion	Horse	2006
28	Sz	A1/4	91	BHS 37	Lower respiratory (severe bronchopneumonia)	Dog	2001
29	Sz	A1/4	1	H70	Upper respiratory	Horse	2000
30	Sz	A1/5	NA	B23 7156	Uterine infection/abortion	Horse	2003
31	Sz	A1/5	47	BHS 28	Lower respiratory	Dog	2000
32	Sz	A1/5	10	BHS 41	Lower respiratory	Dog	2001
33	Sz	NA	16	B24 7132	Upper respiratory	Horse	2004
34	Sz	NA	5	B24 7159	Lymph node abscess	Horse	2004
35	Sz	NA	54	B24 7155	Upper respiratory	Horse	2004
36	Sz	NA	125	B24 7157	Upper respiratory	Horse	2004
37	Sz	NA	17	B24 7158	Upper respiratory	Horse	2004
38	Sz	NA	95	B24 7156	Upper respiratory	Horse	2004
39	Sz	NA	48	B25 4433	Lymph node abscess	Horse	2005
40	Sz	A1/A2	93	B23 7166	Uterine infection/abortion	Horse	2003
41	Sz	A1/D1/5	55	B23 7163	Uterine infection/abortion	Horse	2003
42	Sz	B1/1	8	D40	Lower respiratory	Horse	2000
43	Sz	B1/2	NA	BHS 23	Lower respiratory	Dog	2000
44	Sz	B1/3	2	D14a	Lower respiratory	Horse	2000
45	Sz	B1/4	123	BHS 5	Lower respiratory	Dog	1999
46	Sz	B1/u	86	H52	Upper respiratory	Horse	2000
47	Sz	C1/1	45	D56	Lower respiratory	Horse	2000
48	Sz	C1/2	NA	D88	Upper respiratory	Horse	2000
49	Sz	C1/3	45	H51	Upper respiratory	Horse	2000
50	Sz	C1/u	NA	D44	Lower respiratory	Horse	2000

Table A.4 continued

Lane	Species	PCR type or SeM allele	ST	Strain ID	Disease/Source	Animal	Year
51	Sz	C2/B2/D2	NA	D13	Lower respiratory	Horse	2000
52	Sz	C2/4	6	S13	Skin swab	Dog	2002
53	Sz	C2/u	4	B24 4389	Uterine infection/abortion	Horse	2004
54	Sz	C2/u	149	6618	Uterine infection/abortion	Horse	2003
55	Sz	D2/u	101	B24 2382	Uterine infection/abortion	Horse	2004
56	Sz	D1/D2/4	178	2958	Uterine infection/abortion	Horse	2003
57	Sz	NA	58	B26 Tansey	Nephritis	Horse	2006
58	Sz	NA	96	B25 4763	Lymph node abscess	Horse	2005
59	Sz	NA	46	B25 OO61	Lymph node abscess	Horse	2005
60	Sz	NA	50	B25 5405	Lymph node abscess	Horse	2005
61	Se	SeM-6	NA	7063	Strangles	Horse	2003
62	Se	SeM-7	179	1931	Strangles	Horse	2004
63	Se	SeM-14	179	3156	Strangles	Horse	2004
64	Se	SeM-10	179	7140	Strangles	Horse	2003
65	Se	SeM-12	179	3154	Strangles	Horse	2004
66	Se	NA	NA	VOOO9445	Bastard strangles	Horse	NA
67	Se	SeM-2	179	303	Strangles*	Horse	1999
68	Se	SeM-2	179	SA	Strangles*	Horse	1999
69	Se	SeM-15	179	181063	Strangles*	Horse	1999
70	Sz	D1/1	23	H27	Lower respiratory	Horse	2000
71	Sz	D1/1	23	H30	Lower respiratory	Horse	2000
72	Sz	D1/1	NA	D91	Upper respiratory	Horse	2000
73	Sz	D1/1	23	H29	Lower respiratory	Horse	2000
74	Sz	D1/1	8	H22a	Lower respiratory	Horse	2000
75	Sz	D1/1	NA	D96	Upper respiratory	Horse	2000
76	Sz	D1/3	NA	D10a	Lower respiratory	Horse	2000
77	Sz	D1/5	9	H54	Upper respiratory	Horse	2000
78	Sz	D1/u	NA	P15	Lower respiratory	Horse	2000
79	Sz	D1/u	NA	D19	Lower respiratory	Horse	2000
80	Sz	D1/u	8	D2a	Lower respiratory	Horse	2000
81	Sz	D1/u	8	H49	Upper respiratory	Horse	2000
82	Sz	A1/5	53	B24 7102.2	Lymph node abscess	Horse	2004
83	Sz	A1/5	5	B24 7043	Lymph node abscess	Horse	2004
84	Sz	A1/5	5	B24 7325	Upper respiratory	Horse	2004
85	Sz	A1/u	4	B24 4388	Uterine infection/abortion	Horse	2004
86	Sz	A2/3	14	B24 2519	Uterine infection/abortion	Horse	2004
87	Sz	B1/2	3	D33	Lower respiratory	Horse	2000
88	Sz	B1/2	NA	BHS 2	Lower respiratory	Dog	1999
89	Sz	D2/3	61	H40	Lower respiratory	Horse	2000
90	Sz	D2/u	101	B24 4996	Uterine infection/abortion	Horse	2004
91	Sz	D2/u	56	B23 3322	Uterine infection/abortion	Horse	2003

NA Information not available. Details of many of these isolates are also available on the online MLST database: http://pubmlst.org/szooepidemicus/. Accessed 23.02.09). * Isolates collected from outside the UK.